


Spring 7-24-2013

## Exploring Data Quality of Weigh-In-Motion Systems

Chengxin Dai  
*Portland State University*

Let us know how access to this document benefits you.

Follow this and additional works at: [http://pdxscholar.library.pdx.edu/open\\_access\\_etds](http://pdxscholar.library.pdx.edu/open_access_etds)

 Part of the [Other Civil and Environmental Engineering Commons](#), and the [Transportation Engineering Commons](#)

---

### Recommended Citation

Dai, Chengxin, "Exploring Data Quality of Weigh-In-Motion Systems" (2013). *Dissertations and Theses*. Paper 1018.

[10.15760/etd.1018](https://doi.org/10.15760/etd.1018)

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

# Exploring Data Quality of Weigh-In-Motion Systems

by

Chengxin Dai

A thesis submitted in partial fulfillment of the  
requirements for the degree of

Master of Science  
In  
Civil and Environmental Engineering

Thesis Committee:  
Christopher M. Monsere, Chair  
Robert L. Bertini  
Kristin A. Tufte

Portland State University  
2013

© 2013 Chengxin Dai

## **ABSTRACT**

This research focuses on the data quality control methods for evaluating the performance of Weigh-In-Motion (WIM) systems on Oregon highways. This research identifies and develops a new methodology and algorithm to explore the accuracy of each station's weight and spacing data at a corridor level, and further implements the Statistical Process Control (SPC) method, finite mixture model, axle spacing error rating method, and data flag method in published research to examine the soundness of WIM systems. This research employs the historical WIM data to analyze sensor health and compares the evaluation results of the methods. The results suggest the new triangulation method identified most possible WIM malfunctions that other methods sensed, and this method unprecedentedly monitors the process behavior with controls of time and meteorological variables. The SPC method appeared superior in differentiating between sensor noises and sensor errors or drifts, but it drew wrong conclusions when accurate WIM data reference was absent. The axle spacing error rating method cannot check the essential weight data in special cases, but reliable loop sensor evaluation results were arrived at by employing this multiple linear regression model. The results of the data flag method and the finite mixed model

results were not accurate, thus they could be used as additional tools to complement the data quality evaluation results. Overall, these data quality analysis results are the valuable sources for examining the early detection of system malfunctions, sensor drift, etc., and allow the WIM operators to correct the situation on time before large amounts of measurement are lost.

## DEDICATION

This work is dedicated to my parents, my grandparents and my sister, who have supported and encouraged me through years with love and understanding. I specially dedicate this work to my boyfriend, Yiduo, who has been my great inspiration for my graduate studies and this work. For Nikki, Aunt Jane and Uncle Qiao, thank you very much for helping me to achieve it.

## ACKNOWLEDGEMENTS

I appreciate and am indebted to my advisor, Dr. Christopher M. Monsere, who gave ample encouragement along the way and reassured me that there would indeed be an end. Without his patient guidance and tireless devotion to this work, conducting this research and writing this thesis would not be possible. I would like to thank Dr. Robert L. Bertini and Dr. Kristin A. Tufte for providing excellent suggestions regarding my work and for serving on my committee.

## TABLE OF CONTENTS

ABSTRACT .....	i
DEDICATION .....	iii
ACKNOWLEDGEMENTS.....	iv
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
1.0 INTRODUCTION .....	1
1.1 Problem Statement .....	2
1.2 Research Objectives .....	4
1.3 Thesis Scope.....	5
1.4 Organization .....	6
2.0 BACKGROUND.....	8
2.1 WIM System Operation .....	8
2.2 Data Quality Evaluation Metric .....	11
2.2.1 Weight Data Accuracy Metrics.....	13
2.2.2 Speed and Axle Spacing Accuracy Metric .....	29
2.3 Calibration Drift .....	32
2.4 Temperature .....	34
2.5 Vehicle Dynamics .....	36
2.6 Pavement Roughness.....	38
2.7 Gaps in Current Research.....	39
2.8 Summary of Background Review .....	40
3.0 DATA DESCRIPTION.....	43
3.1 WIM Data.....	43
3.2 Static Weight Data .....	48
3.3 Summary .....	49
4.0 METHODOLOGY.....	50
4.1 Triangulation Method for Sensor Performance Measurement .....	51
4.2 Statistical Process Control Chart.....	56
4.2.1 Subgroup Frequency and Sample Size .....	62
4.2.2 Average and standard deviation chart.....	64
4.3 Data Quality Flag .....	68
4.4 Finite Mixture Distribution .....	70
4.5 Axle Spacing Error Rating Model.....	73
5.0 ANALYSIS AND RESULTS .....	76



5.1	Descriptive Statistics .....	76
5.1.1	Truck Characteristics .....	79
5.2	Weight Data Accuracy Results .....	82
5.2.1	Triangulation Method .....	83
5.2.2	Statistical Process Control .....	86
5.2.3	Data Quality Flag.....	91
5.2.4	Finite Mixture Distribution .....	94
5.3	Axle Spacing Data Accuracy Results.....	98
5.3.1	Triangulation Method .....	98
5.3.2	Statistical Process Control .....	100
5.3.3	Data Quality Flag.....	104
5.3.4	Axle Spacing Error Rating Model .....	106
5.4	Results Comparison.....	109
5.4.1	Rating Criteria.....	109
5.4.2	Weight Accuracy Results.....	112
5.4.3	Axle Spacing Results .....	116
6.0	CONCLUSIONS.....	121
6.1	Summary of Findings .....	121
6.2	Future Work .....	124
7.0	REFERENCES.....	127
8.0	APPENDIX.....	130
	APPENDIX A: Truck Characteristics Analysis .....	130
	APPENDIX C: Mixture Model Temporal Changes .....	143
	APPENDIX D: Weekly Triangulation Plots.....	148

## LIST OF TABLES

Table 1: WIM Systems Study Data Size.....	47
Table 2: Statistical Process Control Chart Types .....	58
Table 3: Average Run Lengths for Various Shift Levels and Run Rules .....	68
Table 4: Coefficient of Variation of Average Axle Weight and Spacing at WIM Stations .....	77
Table 5: Coefficient of Variation of Average Steering Axle Weight at Static Data Collection Sites .....	77
Table 6: WIM Load Cell Scale Data Quality Result Table .....	116
Table 7: WIM Loop Sensor Data Quality Result Table .....	120

## LIST OF FIGURES

Figure 1: Load Cell WIM System in Woodburn Port of Entry in Oregon .....	10
Figure 2: QC/QA Program in North Central Region of FHWA-LTPP .....	17
Figure 3: Variation in Adjustment Factor over Time for Temperature .....	36
Figure 4: Calendar Heat Map of Station 2 Truck Volume.....	45
Figure 5: Truck Volume Calendar Heat Plot at WIM Stations 1, 2, 3, 4, 5 and 6.....	47
Figure 6: Map of Oregon WIM Sites used in Study .....	48
Figure 7: Shewhart Rule Illustration .....	62
Figure 8: Sample Average and Standard Deviation Control Chart.....	66
Figure 9: Sample Average Control Chart using QCC Package of R .....	66
Figure 10: Distribution of GVW in Station 4 .....	72
Figure 11: Histogram of Mixture Distributions in Station 4.....	72
Figure 12: Length of Commonly Used Semi-Trailer Truck .....	78
Figure 13: Time Series Moving Average Plot of Class 9 Trucks at Station 1 .....	82
Figure 14: Steering Axle Weight Analysis in Triangulation Method at Stations 1, 2, and 3 .....	85
Figure 15: Steering Axle Weight Analysis in Triangulation Method at Stations 4, 5 and 6 .....	85
Figure 16: Steering Axle Weight Analysis in Average and Standard Deviation Chart of SPC Method at Station 1 .....	89
Figure 17: Steering Axle Weight Analysis in Average and Standard Deviation Chart of SPC Method at Station 3.....	90
Figure 18: Normalized WIM Weight Sensor Flags at Stations 1, 2 and 3.....	92
Figure 19: Normalized WIM Weight Sensor Flags at Stations 4, 5, and 6.....	93
Figure 20: Two, Three and Four Components in Mixture Model in Component Mean and Proportion Distribution .....	97
Figure 21: Component Means and Proportions at WIM Station 2 .....	97
Figure 22: Drive Tandem Axle Spacing Analysis in Triangulation Method at Stations 1, 2 and 3.....	99
Figure 23: Drive Tandem Axle Spacing Analysis in Triangulation Method at Stations 4, 5 and 6.....	100
Figure 24: Drive Tandem Axle Spacing Analysis in Average and Standard Deviation Chart of SPC Method at Station 1.....	102
Figure 25: Drive Tandem Axle Spacing Analysis in Average and Standard Deviation Chart of SPC Method at Station 3.....	103
Figure 26: Normalized WIM Weight Sensor Flags at Stations 1, 2 and 3.....	105
Figure 27: Normalized WIM Weight Sensor Flags at Stations 4, 5 and 6.....	106
Figure 28: Error Rating of Drive Tandem Axle Spacing.....	108

## 1.0 INTRODUCTION

Weigh-In-Motion (WIM) systems measure the instantaneous dynamic loads applied to a roadway by a moving vehicle without interfering with vehicle movement. The WIM system has been in development since the 1960s (Southgate, 1999). So far, a diversity of technologies has been developed to support WIM systems, including the load cell scale, bending strain scale and piezoelectric weight sensor. The low-cost instrumentation feature and data gathering technology has made the WIM system more affordable and popularized worldwide.

A WIM system produces weight data (e.g., axle weight, group axle weight and gross vehicle weight) and traffic-stream data (e.g., speed, vehicle classification and traffic flow). This system has a variety of applications, mainly in the area of weigh station enforcement. The WIM system is used as a pre-screening tool in the commercial vehicle weight enforcement operation. Furthermore, WIM data have also been used for the purpose of highway capacity design, bridge and pavement design, and planning data assessment for highway improvement programs, the Long Term Pavement Performance (LTPP) Program, the Highway Pavement Monitoring System, and the Mechanistic-Empirical

Pavement Design Guide (Ramachandran et al, 2011; Oman, 2010; Sayyady et al., 2010).

With the increasing WIM applications, the requirement on WIM data accuracy level has been expanding. This study reviews the literature of existing WIM system performance requirements and testing standards, and identifies and develops a new WIM system data quality method. Along with the up-to-date performance methods, new data quality comparisons are explored with historical WIM data from sample Oregon WIM stations. Data mining of each method reveals variations in the data caused by incorrect calibration and sensor failure. The consequences of each data quality methods are compared in performance-based rating criteria. The strength and weakness of data quality approaches are analyzed.

## **1.1 Problem Statement**

WIM systems provide a great deal of valuable data in an efficient manner. Many state agencies have been using WIM systems for weigh station enforcement, and this requires a high level of data quality (Southgate, 1999). The WIM data are currently not accurate enough to support the weight enforcement (Nichols & Bullock, 2004). Some state agencies do not have time or data resources

to examine quality assurance, and most WIM sensor vendors do not include details for quality control in reports (Southgate, 1999). Therefore, the need for quality assurances of WIM data became apparent. A level of data accuracy must be defined in the quality control process to improve the WIM system performance.

The WIM data accuracy has been controversial. A WIM system is designed to estimate a static vehicle weight based on measurements of the dynamic forces load on the sensors, but the dynamic weight output can be significantly different to the actual axle weight measured from a static scale. The WIM weight estimator is related with a broad variety of factors, including WIM sensor technology and health, system calibration and monitoring, pavement profile, vehicle characteristics and dynamics, and temperature and precipitation. Vehicle characteristics specifically include suspension characteristics, truck center of gravity, tire width and pressure, and frame length (Southgate, 1999; Ramachandran et al, 2011; Han et al., 1995).

System calibration ensures a collection of good quality weight data by correcting data to closely approximate the true force applied to the sensors (Dahlin, 1992; Ramachandran et al, 2011). The calibration factor is closely related

to the WIM systems' application: when calibration is too low, the low recorded axle loads result in insufficient fatigue calculations and thin pavement thickness designs. On the contrary, a calibration factor that is too high causes excessive pavement thickness designs and excessive expenditure (Southgate, 1999).

Variations from established properties in the data may be on account of a malfunctioning system, calibration drift and truck dynamics. As to the truck dynamics, when WIM systems are in good operation, there will be a gap between dynamic and static weight. The dynamic axle force has only 10 to 30 percents RMS amplitudes of the static axle loads of heavy goods vehicles (Cebon & Winklet, 1991). The tire force of a dynamic vehicle caused by oscillation partially accounts for the weight difference between dynamic weight data and static weight data (Cebon & Winklet, 1991).

## **1.2 Research Objectives**

The measurement accuracy of WIM technology is of prime importance. A well-defined data quality control method helps evaluate data to meet requirements on the applications of bridge and pavement design, system performance and planning data.

The objective of this research is to explore multiple criteria that the Oregon Department of Transportation (ODOT) can use to assess WIM data quality. To evaluate performance, this research: (1) reviews current research and methodologies to study WIM data quality to identify appropriate research methods; (2) identifies and develops a new corridor-level data quality control metric; (3) applies the selected data quality metrics to the historical Oregon WIM data to evaluate the health of weight scale and axle spacing sensors, respectively; (4) compares the findings of diverse data quality metrics in weekly erroneous data and provides recommendations for the methods application.

### **1.3 Thesis Scope**

This research focuses on WIM data accuracy by examining Class 9 truck data in five data quality methods. WIM quality control procedures have different levels. On a broad scale, the LTPP Quality Assurance Program commonly guides WIM quality control in many applications. It identifies potential errors in WIM data and summarizes the datasets in a series of graphs to determine unusual occurrences in the submitted traffic data. It works for studies not intended for submission to the LTPP program (McCall & Vodrazka, 1997). The Vehicle Travel Information System Software (VTRIS) is another method to validate vehicle



travel characteristics by WIM site characteristics adjustment, and incorporates these validated data into permanent WIM database (McCall & Vodrazka, 1997). On a small scale, there are a number of state- and project-specific traffic data quality control requirements, including a truck re-identification project (Monsere, 2011), and truck axle spectra development for use in pavement design and analysis (Elkins & Higgins, 2008). This is a state-level WIM data study that explores the WIM data quality metrics.

#### **1.4 Organization**

This thesis is organized as follows: Chapter 2 provides a review of WIM data quality methods in weight and axle spacing data accuracy metrics to examine erroneous data resulting from a malfunctioning system or calibration drift. Chapter 3 introduces the selection of study locations and historical WIM data sources in the search. Chapter 4 discusses a WIM data quality case study which adopted four research methods for steering axle weight and drive tandem axle spacing. Chapter 5 analyzes the results of four data quality methods in weight and spacing, integrates all methods' results by shared rating criteria and compares each method's results. A summary of comparison results by weight and spacing data type is provided at the end of the chapter. Chapter 6 provides a

summary of conclusions and recommendations for future research. The summarized findings are closely related to WIM system data quality methods in state-level WIM system practices. Insight regarding future predictions of data quality control requirements is provided.

## **2.0 BACKGROUND**

This chapter provides a literature review of WIM systems data quality methods and an understanding of influential factors affecting WIM systems data quality. Here, the reader will find a description of WIM data quality methods used to evaluate the system operation, guidance of procedures to apply them into practice, and their advantages and disadvantages in diverse contexts.

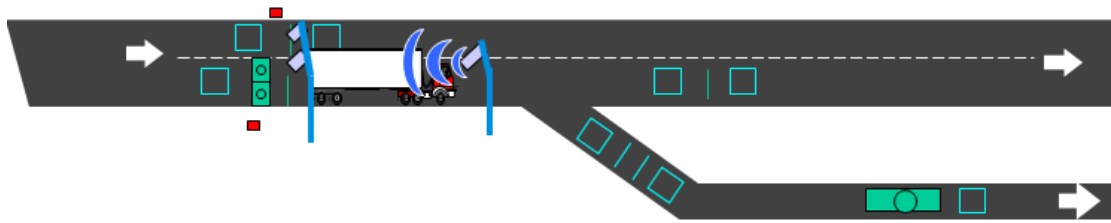
### **2.1 WIM System Operation**

According to the American Society of Civil Engineers 2009 Report Card, highway infrastructure continues to degrade due to the absence of government funding for rehabilitation, maintenance and reconstruction. The WIM system in Oregon, also known as the Green Light WIM System, is a preclearance system that weighs vehicles and their cargo as each vehicle crosses the scales while in motion (Cook, 2011). The WIM system measures essential data including distribution of axle loads, equivalent standard axle loads, classification of trucks, and percentage of overloaded heavy vehicles.

Oregon's prescreening program adopts truck weigh station devices at 22 locations and allows trucks equipped with a transponder to bypass the weigh

station at highway speeds. ODOT located these truck weigh station devices along the mainline of freeways and upstream from on-ramps. Oregon's truck weigh station devices are the typical type of load cell WIM system. The mainline has two in-line single load cell scales for weight measurement, a pair of inductive loop detectors for detecting vehicles and activating a nearby computer for data transfer, and a height detector for height enforcement, as shown in Figure 1. The loop sensors are located upstream and downstream of the single load cell for axle spacing measurement. This load cell WIM system utilizes two single load cells to detect and weigh the right and left sides of an axle simultaneously.





**Figure 1: Load Cell WIM System in Woodburn Port of Entry in Oregon**

When a WIM system is in operation, the radio-frequency identification (RFID) reader activates the vehicle's transponder and sends a unique acquisition signal to the computer. The WIM station receives all vehicle data and checks state records such as registration, weight declaration, tax status and safety inspections. Then it sends a signal to communicate with the transponder-equipped vehicles. Within a second, the driver is signaled on the transponder with a green light to go or a red light to stop and pull over. The non-compliant vehicles or the vehicles not equipped with a transponder are pulled onto the ramp into a weighing facility and re-enter the interstate lanes after examination (Cook, 2011).

The load sensor is a 6-foot platform scale built in pavement allowing each wheel set to be weighed individually, axle by axle. The scale mechanism - load cell sensor - has load transfer torque tubes which effectively transfer tire loading on the weighing surface to the load cell. At present, the single load cell WIM scale is the highest documented accuracy of all WIM sensors. It is also

maintainable and has a long service life. A properly installed and calibrated single load cell WIM system can provide Gross Vehicle Weight (GVW) within 6 percent of the actual vehicle weight accuracy for 95 percent of the measured trucks (Scholz, 2012).

At a single load cell WIM site, the speed is calculated using the traveling time  $Time_{Loop}$  and fixed distance  $Distance_{Loop}$  between inductive loop sensors, as shown in Eq. 1. The speed is used to calculate the axle spacing. The first inductive loop sensor measures the time difference hit by two consecutive axles  $Time_{axle\ sensor}$ . With truck speed and time of passing consecutive axles, axle spacing can be calculated in Eq. 2 (Nichols & Bullock, 2005).

$$Speed_{space\ mean} = \frac{Distance_{Loop}}{Time_{Loop}} \quad \text{Eq. 1}$$

$$Distance_{axle} = Time_{axle\ sensor} * Speed_{space\ mean} \quad \text{Eq. 2}$$

## 2.2 Data Quality Evaluation Metric

WIM data quality mainly includes weight and axle spacing data quality, in which weight data are key elements of WIM systems. Previous research efforts have identified weight accuracy metrics by applying them to the most predominant truck type - FHWA Class 9 vehicle fleet.

An ideal WIM data quality check would allow little variability, but it would primarily detect outliers or disturbances that are very likely outside the expected average. Researchers have proposed a variety of data quality metric in steering axle weight (Southgate, 1999; Ott & Papagiannakis, 1996; Grundy et al., 2002); gross vehicle weight (Nichols & Bullock, 2006; Grundy et al., 2002); drive tandem axle spacing (Nichols & Bullock, 2006); drive tandem axles weight (Grundy et al., 2002); spacing between front axle and lead driving axle (Grundy et al., 2002; Southgate, 1999); the left and right of front wheel weight residual (Nichols & Bullock, 2006); and single axle loads (Kweon and Cottrel, 2011). Some research derived sensor failure proportions to identify calibration drift (Nichols & Bullock, 2004; Prozzi & Hong, 2007). Agencies utilize sets of these metrics in WIM quality assurance programs to ensure truth-in-data before uploading to a database (Han et al, 1995; Pelphrey et al., 2008).

Data quality control methods give insight into the source of problems that might cause erroneous data records. If errors occasionally occur, the problem is probably a terminal connection that is affected by moisture or temperature. If errors are continuously detected for a while, the problem is probably an equipment malfunction (Nichols & Bullock, 2004).

Most research mentioned above examines the WIM data quality using the Class 11 five-axle, tractor-semitrailer truck fleet, which is the most predominant truck type on the U.S. national highways. A Class 11 in the ODOT classification scheme is a Class 9 in the FHWA classification scheme. Its acronym “3-S2” indicates a three-axle tractor pulling a trailer with two axles, the most predominant commercial vehicle class on U.S. highways. Therefore, it is commonly used for analyzing data accuracy by many agencies.

### **2.2.1 Weight Data Accuracy Metrics**

Weight accuracy is a key aspect of WIM data application. Without accurate weight data, a WIM system is only an expensive vehicle classification system and a traffic data counting system. Most metrics are related with gross vehicle weight, steering axle weight, and steering axle wheel weights, and these weight data accuracy metrics are sound and straightforward. The GVW metrics are based primarily on visual analysis of histograms and kernel density estimates, which are relatively subjective.



### **2.2.1.1 Traditional Weight Accuracy Approach**

Different WIM application requires different data accuracy. The North Central Region of FHWA-LTPP released a weight data quality control program, as shown in the flowchart Figure 2. The chart shows the quality control process and how operational techniques are employed to meet the data requirements. The widespread use of weight data made weight data quality evaluation critical.

In 1990, American Society for Testing and Materials (ASTM) specification E1318-02 outlined a calibration approach for WIM data quality assurance. Test traffic stream trucks are designed to be weighed at a static weight station before entering a WIM station. The static scale was located three miles upstream from the permanent WIM site. Instrument trucks are regulated to be in different configurations to reduce the calibration bias of a particular vehicle class and its response to pavement profile (Nichols et al., 2009). These dedicated trucks make repeated trips over the scale. The objective load variables may be chosen as steering axle load, drive tandem weight, trailer tandem weight, and GVW.

This traditional calibration process of WIM devices has been argued as an unpractical process for its limitations and problems (Dahlin, 1992; Southgate, 1999). It requires either a static weight station in the immediate vicinity or portable static scales, and this method solely relies on test trucks. To identify

designed trucks in a normal stream at the WIM site with data collected at the static loading meter station is tough. Another important point was that WIM devices often produce various dynamic weights for each test truck because of vehicle dynamics. Test trucks are far less representative of the weights of all trucks in the traffic system (Dahlin, 1992). Lastly, it requires multiple passes over the sensors to determine calibration factors. For large networks of WIM devices, the traditional calibration method is costly to operate on a regular basis. Rental trucks or those owned by the agency are hard to schedule because of turn-around sites and the time between runs (Southgate, 1999).

Using dedicated trucks for calibration efforts has some advantages and disadvantages (Southgate, 1999). For the benefits of test trucks, it cost-effectively takes advantage of existing permanent static scales; the trucks are always available, having no rental cost or administrative accounting costs, and characteristics are obtained such as suspension type and possible runs for braking; and the knowledge of drivers on the designed trucks becomes an asset when attempting to provide good correlation efforts. However, the operational disadvantages for using certain trucks for calibration efforts are fleet schedule and operation expense. Crew cost could be high if engaging in calibrating the

scale for a number of days. Also, it limits the number of calibration runs (Southgate, 1999).

#### **2.2.1.2 GVW and Steering Axle Load**

Dahlin (1992) proposed a method to determine the validity of weight data over time. The visual evaluation specially focused the distribution of GVW, steering axle weight, and accumulated flexible Equivalent Single Axle Load (ESAL) factor in turn with a series of criteria. Dahlin (1992) found that GVW mostly has two peaks: one for empty trucks weighing between 28,000 and 32,000 pounds and the other for laden trucks ranging between 70,000 and 80,000 pounds. An instrument malfunctioning WIM station is reflected as not having the classic bimodal GVW distribution or the expected patterns. That suggested that the steering axle weight should fall in a certain range of values corresponding to the GVW. If average steering axle loads fall out of this range, recalibration is needed. Additionally, the number of ESAL accumulated from 3-S2 trucks in each lane should have a consistent pattern in each week.

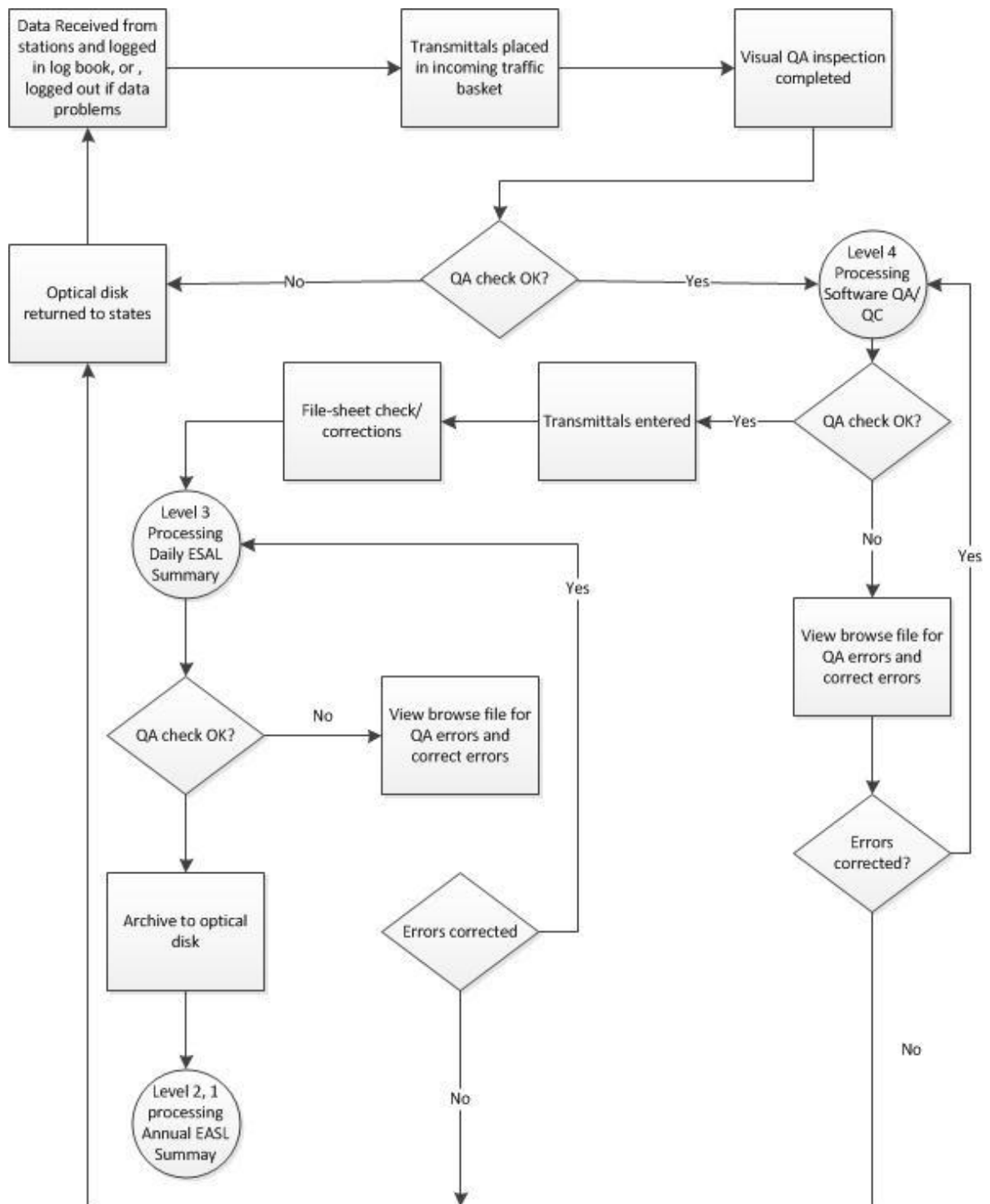


Figure 2: QC/QA Program in North Central Region of FHWA-LTPP

Han et al. (1995) questioned Dahlin's study because it requires many samples and extensive experience with WIM systems to interpret the plotted results. At some WIM sites with low traffic volume, frequency distribution patterns cannot reasonably be in the binomial distribution. Also, the distribution check requires bin width selection, which is overly related to the number of trucks in each bin. A wider bin is likely to distort calibration drift. Most importantly, Dahline's research cannot provide quantified judgment on calibration drift, which is often urgently needed to adjust data distorted by the drift in practice. Han et al. (1995) expanded Dahlin's approach into a more objective method that automated the checking process to determine system quality. The modified approach considered two attributes of variations of steering axle loads in the 3-S2 fleet - one was axle loads of a fleet of vehicles that were weighted statistically and the other was the variation caused by vehicle dynamics. Chance variation of weight measurement of WIM devices was not studied in Han's study, although it can cause steering axle load variation as well.

Nichols and Bullock (2004) also rebut Dahlin's method. They stated that the method required for this metric is burdensome and "not intuitive for adjusting the calibration to obtain accurate axle spacing." Furthermore, this metric requires

accurate speed calibration to obtain accurate axle spacing, but accurate individual speed data are not guaranteed. In addition, the 3-S2 GVW distribution method is based on the assumption that the proportion of loaded and unloaded trucks would remain constant over time, but this condition is seldom met.

GVW distribution of FHWA Class 3-S2 vehicles are usually used for assessing WIM accuracy by visual interpretation of frequency histograms. Another data quality assurance method using gross vehicle weight analysis was developed by Nichols and Cetin (2007) using mixed distribution models. This method overcomes the disadvantages of the previous study as it does not need the data to be binned. Nichols and Cetin (2007) implemented finite mixture models to fit a combination of two or three normal distributions to the overall GVW distribution. The bimodal or trimodal GVW distribution has prominent peaks in the loaded and unloaded ranges. Mixture models accurately identify peak group mean, covariance and proportion of traffic belonging to those peaks for ongoing monitoring purposes. This study used historical Indiana WIM data to illustrate this numerical characterization analysis method and benefits. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to determine a suitable number of components of mixture models.

Larsen & McDonnell (1999) examined the data quality of quartz-piezoelectric WIM sensors in Connecticut through support from the FHWA-LTPP program. The sensor survivability and accuracy can be determined from automated GVW distribution analysis. The weekly GVW distributions of all vehicle types by lane of station showed the shift of peaks that corresponded with calibrations, and the loaded vehicle characteristics that were to monitor the system over time. The GVW distribution also determines the likely range of unloaded vehicles and loaded vehicles, respectively.

Ott and Papagiannakis (1996) developed an innovative WIM data quality assurance method by analyzing the variation in measurements of the steering axle load variation of FHWA Class 9 trucks. Confidence-interval limits of these measurements are calculated by using historic static loads adjusted for the effect of air resistance and their combined variation from two sources: the variation within a five-axle semitrailer vehicle fleet and the variation due to axle dynamics at a WIM station. The variation within a 3-S2 fleet vehicle was determined by analyzing historical static axle load data from two resources, the Truck Weight Study database maintained by FHWA and the state of Washington's database. The variation of static steering axle loads within a fleet of 3-S2 trucks was small

with coefficients of variation less than 12 percent. The axle load variation caused by vehicle dynamics was determined through dynamic vehicle simulations of a vehicle model VESYM and pavement roughness on site. The model VESYM was calibrated by comparing predicted axle acceleration and frame location on-truck with measured data at locations using accelerometers. The study found the variation of dynamic axle load increases with a high vehicle speed. For the range of roughness on flexible highway pavement, the coefficient of variation in dynamic axle loads was experimentally found to be as high as 20 percent (Ott & Papagiannakis, 1990). The extent of steering axle load variation was reduced by considering fleet subgroups by their GVW. Additionally, Ott and Papagiannakis (1996) analyzed GVW data from the Truck Weight Study (TWS) database, selecting 976 WIM sites, which represent a rough sampling ratio of 15 percent. Regarding the TWS database, roughly 80 percent GVW distributions are bimodal. The GVW frequency distribution pattern is summarized and divided into four types:

1. Bimodal with two distinct peaks representing unloaded and loaded trucks
2. A dominant peak for empty trucks



3. A dominant peak for loaded trucks
4. Multimodal distributions with more than two or three distinct peaks

Ott and Papagiannakis (1996) and Dahlin's (1992) GVW distribution analyses are prior to Nichols and Cetin (2007) GVW distribution study, which resulted with classic trimodal distribution. It is hypothesized that these vehicles are carrying partial loads or full loads of low-mass cargo (Nichols & Cetin, 2007).

Grundy et al. (2002) developed a method to accommodate time dependent drift of WIM data. This research considered the system variation (known as chance variation) as a function of characteristics of trailer and tractor, and the controllable variation (known as assignable variation) as a function of environmental effects on the measurement system. A real-time adjustment factor is calculated by categorical regression analysis. The analysis significantly reduced the variation in the mean estimates of group axle weight and GVW, and compensates for daily and hourly dependent calibration drift. Variables of regression analysis are steering axle weight, GVW, steering-to-leading drive tandem axle spacing, drive tandem axle spacing, week of the year, and hour of the day. The study found consistent adjustment factors by hours of a day and

week of a year as to steering axle weight when axle spacing and drive axle mass are taken into account. Additionally, the increasing legal weight limit for trucks inevitably leads to increases in GVW, but not for individual axle weight (Grundy et al., 2002).

Sayyady et al. (2009) reviewed WIM volume and weight data in a quality control procedure. The proposed approach only reviewed the completeness and anomalies of WIM volume and truck axle load data in a quality control procedure and did not include the data accuracy analysis.

Southgate (1999) inspected the variability of static weight data of Kentucky's WIM system. Southgate showed the relationship between steering axle weight and gross vehicle weight is not very precise, but there is a relationship between steering axle weight and the spacing between steering axle and the lead drive axle. He proposed an upper bound limit on the relationship using the manufacturer's maximum allowable load and a lower bound limit. However, this metric is not practical to implement because it does not have a clearly defined model making sound data quality conclusions. He developed a new metric - the "tire load ratio" - which is the load per tire on the steering axle divided by load per tire on the drive tandem as shown in Eq. 3. The relationship

between the data quality metric “tire load ratio” and total tractor load were found through this metric analysis. The results revealed that tire load ratio is about 3.5 for an empty truck, to 1.1 for a legally loaded truck, and had a curve pattern. The results are from unpublished research so the figures are not available.

$$\text{tire load ratio} = \frac{F_{\text{steering axle tire}}/n_{\text{steering axle tire}}}{F_{\text{drive tandem tire}}/n_{\text{drive tandem tire}}} \quad \text{Eq. 3}$$

System-related quality issues can be minor or major calibration drifts or system malfunction. A slight drift can affect system integrity nearly as seriously as minor and major drifts because a fourth-power relation between weight and ESAL magnifies the slight drift. Southgate (1999) further concluded that three reasons could cause the WIM data to be less than the true axle load:

1. Calibration factors for the WIM scale are too low;
2. Only one side of the truck is passing over the WIM scale, or a portion of the tires on that one side are passing over the scale when a driver is weaving; and
3. The truck may be accelerating or climbing a grade.

### **2.2.1.3 Left-Right Steering Axle Weight Residual**

In 1987 Izadmehr and Lee found that some popular WIM systems such as load cell and bending plate weighed the left and right wheel separately, and they examined the weight difference between left and right wheels of the steering axle. In 2004 Nichols and Bullock identified a derived metric to evaluate the weight accuracy of WIM data with considerations of vehicle dynamic and rough pavement. The left-right residual  $\Delta_{LR}$  is calculated as the difference between the left-right wheel weights as a percentage of the total steering wheels weight as shown in Eq. 4, where  $F_L$  is the left front wheel weight, and  $F_R$  is the right front wheel weight. The proposed left-right steering axle load differential is to monitor the calibration drift over time, and the metric is applied Statistical Process Control (SPC) charts to statistically identify significant variations that sensors are exhibiting.

$$\Delta_{LR} = \frac{(F_L - F_R)}{(F_L + F_R)} \times 100 \quad \text{Eq. 4}$$

With this metric, SPC can provide an evaluation procedure for discerning between noise in the data and process drift. Nichols and Bulluck (2004) found the left-right steering axle weight residual is most robust on the steering axle load. Weight accuracy in left-right wheel weight residual is more reliable than the steering axle weight metric. In 2009, Nichols, Bulluck and Schnneider further

expanded the study to the left and right residual metric to other four axles of FHWA Class 9 vehicle fleets and explored the relation between the metric and cross-slope. The left-right residual is a function of cross-slope on the roadway, the engine center of gravity, the amount of fuel in the left - right fuel tanks, the weight and location of people and items in the cab, and the weight and location of the load on the trailer. The average weight distribution of a truck in a station was assumed to be consistent over time under good WIM operation. For a large population of FHWA class 3-S2 vehicles, the effects of truck characteristics are insignificant, and the average calculation can average out any distribution deviations caused by individual truck's characteristics. The left-right weight residual of front wheels was assumed to be associated with cross-slope (Nichols et al., 2009). Typical cross-slopes of interstate highways are around 3 percent. An approximate cross-slope effect on left-right distribution can be estimated by the vehicle's center of gravity. The load of the 3-S2 steering axle is mainly from the weight of the engine and chassis, which is evenly distributed (Nichols et al., 2009). When a truck is driving in a passing lane of a typical two-lane highway, the left wheel will weigh more than the right wheel, yielding a positive residual and vice versa. Nichols and Bullock (2005) calculated monthly average left-right

weight residual to monitor and detect the differential drift in the wheel sensors. They found the left-right residual of a 3-S2 fleet population effectively identified wheel sensors that may be drifted and whether they are read too heavy or too light.

The metrics of steering axle left-right residual and drive tandem axle spacing are consistent for FHWA Class 9 vehicle fleets. Nichols and Bullock (2004) implemented the SPC  $\bar{X}$  chart, s chart and p chart with these data quality metrics to control the WIM data quality. According to sales data and data from sites with good calibration, the spacing control limits were calculated as the same statewide, but these spacing control limits would be better to calculate using the accurate data at WIM sites due to the diverse station characteristics. Error control charts, also known as a fraction defective type p-chart, include errors generated due to abnormal sensor output. The control limit is based on the proportion of errors to the total number of vehicles.

The left and right wheel weights are critical to determine if there are a high number of invalid measurements. When the difference between the two exceeds 40 percent, the measurement for that vehicle is classified as invalid. The large imbalance may be caused by

1. A truck not driving in the middle of the lane
2. Bouncing by empty trucks
3. Empty van trailers in heavy cross winds
4. Bad calibration factor
5. A system malfunction

A program on the WIM site from the host PC determines invalid measurements. If the extent of invalid measurements is still suspicious after reviewing the site characteristics, a check should be made on calibration parameters. Bouncing and cross winds have been causing some percentages of invalid measurements. The independence of left-right wheel weight residual metric measures vehicle dynamic and identifies WIM sites with pavement roughness problems, as discussed later in Section 2.6.

There are two problems with using the SPC method in this study. If we manually calculate the control chart limit parameters, the existing data are not qualified because only valid and accurate data should be used to establish control chart limits. The most appropriate data to base the control limits are the two weeks after a site is calibrated when it is unlikely to have WIM sensor drift or sensor failure. In Oregon's WIM system, the in-control good data are

unknown. Another problem is that the left-right steering axle weight difference is found to be among the most robust metrics. In this study, the left and right steering axle weight data are not available.

In summary, the SPC quality control method is to adjust WIM data to correct seasonal and diurnal drift based upon consistence of left-right steering axle weight and drive tandem axle spacing. It prioritizes maintenance to more effectively repair sensors and determine calibration factors (Nichols & Bullock, 2004).

### **2.2.2 Speed and Axle Spacing Accuracy Metric**

WIM speed data are used to determine axle spacing in bridge formula compliance and derive weight calibration factors in highway improvement program such as Mechanistic-Empirical Pavement Design Guide (MEPDG). Low-quality speed data would directly calculate axle spacing data at a non-acceptable level, which misclassifies vehicle type, generates erroneous weight calibration factors, and leads to improper decisions on bridge formula compliance. A questionable vehicle classification (vehicle classification 0) is a prompt to review traffic speed distributions for that period of time in the on-site speed report. Reasons for low-quality speed data are various, including failure of



periodic speed validation, adjusting sensor distance and recalibrating after sensor replacement (Nichols & Bullock, 2005). Additionally, long traffic congestion could be clearly detected by speed patterns and traffic volumes.

Nichols and Bullock (2006) explored the relationship of speed and axle spacing by measuring the truck speed with a laser gun on WIM sites. It was concluded that the lane with 15 percent higher average speed has 15 percent higher average drive tandem axle spacing and greater unclassified five-axle similar tractor vehicles. When all axles spacing in a lane is overestimated, the axles spacing is beyond the predefined thresholds in the truck classification. If recalibrated, unclassified vehicle count decreases. For example, a 10 percent error increase in calibration distance between two sensors can cause a corresponding 10 percent speed increase due to error, and finally cause a 10 percent rise in vehicle axle spacing. Nichols and Bullock (2004) also applies to drive tandem axle spacing metric to the SPC method. The results showed that SPC could be very effective for detecting when axle sensors are drifting or failing.

The accuracy of axle spacing is assessed by examining the drive tandem axle spacing of the 3-S2 truck population. Inductive loop sensors produce speed to determine the axle spacing. The time between consecutive axles crossing the

measurement is used to calculate axle spacing. Axle spacing data are not only used to classify the vehicles and check their compliance with bridge and pavement specifications, but also are commonly used in weight calibration when a WIM system cannot be analyzed in conjunction with a static scale. Therefore, the accuracy of axle spacing measurements is one of the prerequisites for accurate and stable loading information.

Evaluating the error in the drive tandem axle spacing can assess the soundness of the WIM data (Slavik & Wet, 2012). The total variation includes the true spread of drive tandem axle spacing and the axle spacing measurement error, both of which are assumed to be distributed normally. The error rating (ERas23) in the measurement of drive tandem axle spacing diagnoses the health of WIM devices. An excessive error rating over 7 percent suggests a malfunctioning system. According to the WIM performance in South Africa, the drive tandem axle spacing should be close to 4.43 feet in good operating condition (Slavik & Wet, 2012). From a truck manufactures's specifications review, the spread of drive axle spacing around mean is  $\pm 3$  percent. Slavik and Wet (2012) also suggest this model is most appropriate for the six or seven axle trucks. But the Slavik and Wet study has an issue when applied in other practices.

From the study, it can be inferred that the selected WIM station installed a pair of Class I piezoelectric sensors, but some agencies might use other WIM technologies such as load cell system. Therefore, when applied to other WIM systems, the error rating method merely evaluates loop sensor health rather than the entire WIM system.

### **2.3 Calibration Drift**

The periodic WIM data accuracy studies are an imperative calibration procedure to examine the performance of WIM systems. The calibration procedure ensures that weight estimates are as close to the actual weights as possible (McCall & Vodrazka, 1997). In general, an initial calibration ensures that sensors correctly report vertical force with considerations of weight variation; the following effort is to consistently monitor to identify significant variation of sensor performance and then to recalibrate the sensor. A WIM system is operational after initial field adjustment and absolute calibration. The initial absolute reference can be sampled into subgroups to determine centerline and control limits for the construction of a control chart.

In WIM systems, electronic weight sensors drift as environmental factors such as temperature change. The drift causes weight estimates to be off from

their actual weights. WIM scale manufacturers provide procedures to adjust the calibration factors based on the moving average of a certain number of 3-S2 trucks (Southgate, 1999). Obtaining a moving average of steering axle loads of 3-S2 trucks is a common calibration strategy.

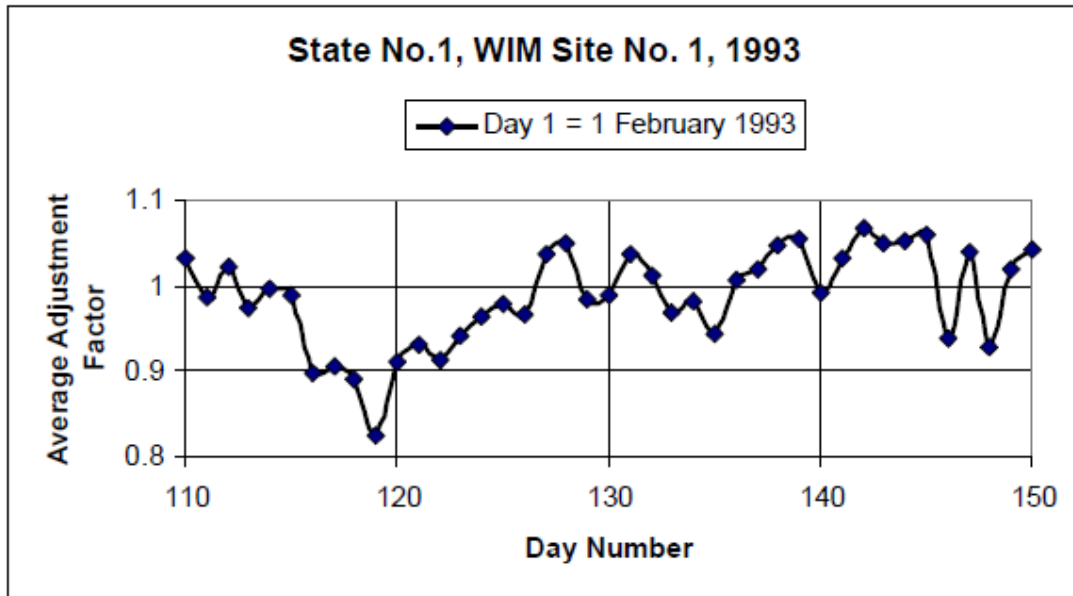
The calibration drift often occurs too slowly to be detected in time. Existing WIM data quality methods cannot effectively detect a small and consistent calibration drift; thus a calibration drift happens usually faster than a calibration interval (Nichols et al., 2009). When sensors drift without any obvious system changes or indications, people won't be aware that a problem exists until weight testing results have been compromised. Hence, if a site has no sign of calibration drift after a period of time, there is usually a need to examine the performance (Southgate, 1999). A long-term calibration drift can be caused by a slow degradation of sensor properties over a long period of time; thus generally the initial calibration cannot discover long-term drift until months later. It can also attribute to the slow changes in signal output. Manual audits of plots that qualitatively assess suitability of data can effectively identify the calibration drift (Southgate, 1999).

Papagiannakis et al. (1996) explored two methods for evaluating and calibrating WIM systems. The first method uses test trucks in the vehicle simulation model VESYM, and the second method compares static and dynamic axle loads of vehicles through automatic vehicle identification. The first method was not able to provide the dynamic axle load over sensor, and was merely used for establishing variation range of a particular WIM site. The range is calibrated by comparing predicted axle acceleration and frame location. In the second method, errors were calculated as the percentage difference between WIM and static loads for individual axles and axle groups.

## **2.4 Temperature**

The widely used weight sensors - bending plates, piezoelectric and load cell - respond to wheel loads with diverse weight accuracy. Sensors' responses to natural temperature changes are similar (Nichols, 2005). Scale manufacturers understand that temperature affects electronic equipment, and thereby the scale readings have been tested to incorporate temperature effects. Manufacturers also provide a temperature adjustment procedure for adjusting the electronic signal from the sensor. From sensor vendors' technical data, the sensor operating temperature ranges from -40°F to 160°F. However, the temperature effect on the

stiffness of flexible pavements is not calibrated, but is also noteworthy (Southgate, 1999). As the pavement temperature increases, the pavement stiffness decreases and tire load causes pavement material to be squeezed. The force that should return to support the WIM sensor is “absorbed” by flexible pavement material (Southgate, 1999). Thus, the energy is diverted to other directions, and the vertical force is less than expected. Contrarily, if pavement temperature is lower, the stiffness increases and vertical energy being returned to support the WIM sensor is higher. Figure 3 shows the daily adjustment factor varies for consecutive days of data in the Southgate study. This is caused by the variation in pavement stiffness due to temperature of the pavement beneath the WIM system, and the variation is separated from temperature effects upon the signal from the sensor. The additional effect is a function of supporting material under the scale.



**Figure 3: Variation in Adjustment Factor over Time for Temperature**

A frequent sensor error can be attributed to sensor technologies, but the occasional equipment errors might be because moisture and temperature affect a terminal connection. If consistent errors occur throughout a day, it is probably due to a malfunctioning loop board or bad terminal connection in nature (McCall & Vodrazka, 1997).

## **2.5 Vehicle Dynamics**

For studying the vehicle dynamics influence, force transducers and accelerometers have been installed on instrument vehicles to measure vehicle dynamics while traveling at speed (Nichols and Bullock, 2006). Hundreds of

instrument vehicles were weighed at both static stations and WIM stations in Indiana. It has been found that the variance in axle weight errors increase as vehicle dynamics increase. Other research (Southgate, 1999) draw the conclusion that normal dynamics of suspension can provide variations in recorded axle loads as much as 15 percent.

Blab and Jacob (2000) provided a solution for dynamics oscillation that is to implement load configuration measurement with a multiple sensors array of 24 piezo-ceramic strip sensors spaced non-uniformly. Instrument vehicles were randomly selected and weighed on static stations located upstream of the testing WIM site. It was found that vehicle dynamics' response to pavement profile is spatially repeatable (Blab & Jacob, 2000). That is, the dynamic loads of heavy trucks clustered at particular positions along the road. As a result, a vehicle dynamics test considerably extended the validity of the results (Blab & Jacob, 2000).

Determining the departures from the established distribution at WIM sites is subjective. In contrast, both the SPC method and the error rating method have objective statistical value to evaluate the erroneous data. Furthermore, an innovative "multiple WIM sensor" configuration allows multiple snapshots of



vehicle forces, and thereby enhances the accuracy of static weight estimates (Blab & Jacob, 2000). The multiple-sensor configuration cannot solve the “tailgating” problem. That is, on heavily trafficked highways, when a moving truck is weighing on site the shock of hitting the sensor induced by the truck’s front axle can be taken as an axle of the last passing truck.

## **2.6 Pavement Roughness**

Unsmooth pavement causes a wide range of vehicle dynamics (Southgate, 1999). Nichols and Bullock (2005) evaluated the level of pavement smoothness with the drive tandem axle group weight residual performance measure. This test was conducted on a WIM station located on a two-lane highway, with one lane of travel in each direction and no noticeable grade. The magnitudes of standard deviations of daily drive tandem axle weight in two lanes are similar, but magnitudes of mean are off by 1.6 kips. The vehicle counts, speeds, and front axle weights in two lanes at one WIM site didn’t reveal any tendency to account for this difference. Thus, the difference was accounted for by different levels of pavement roughness. Multiple types of WIM sensor configurations may be able to solve this problem because sensors receive the same pavement smoothness and pavement composition

## 2.7 Gaps in Current Research

Most researchers have analyzed dynamic weight data in conjunction with static weight data for validating models and verifying experimental assumptions. Variation of dynamic axle loads is a function of vehicle speed and pavement roughness, load configuration and suspension response. However, the extent of variation in dynamic axle loads cannot be completely considered due to the failure to access pavement roughness data, load configuration data and suspension response data. Load transfer due to drive-axle torque is not included in this study either because the effect of driving axle torque could not be easily measured with the clutch engaged and disengaged.

A particular error arises during traffic congestion. If a second truck is tailgating another truck, the steering axle load of the second truck would be recorded as an axle load for the first truck. Then the second axle load of the second truck may be recorded as the steering axle load for the first truck. This situation requires a “Time-Out Factor” WIM setting (Southgate, 1999). Real-time monitoring of a WIM setting is not accessible. An obvious increase in a type of vehicle population usually means a problem of ghost axles being read by the

scale. An abnormal increase in Class 1 report results from erroneous low speed and shortened axle spacing caused by loop error.

Several critical data are not available for this study, including the calibration schedule, left and right vehicle wheel weight data, the corresponding static vehicle weight data, and the vehicle misclassified data (error count). Static vehicle weight is commonly used for WIM calibration and data quality control. The calibration schedule is used for locating the accurate WIM data that consider reference data for most data quality analysis. The misclassified vehicle data can give error count and error proportion for the use of non-conforming statistical control charts (p chart) for monitoring the proportion of the error count relative to the total vehicle count each day. The WIM warning error is an attribute of that vehicle record which could be axle on sensor too long, tailgating, upstream loop only, etc.

## **2.8 Summary of Background Review**

The review presented here informed the research investigation. The chapter reviews the evaluation methods of WIM data quality in both weight and axle spacing measurements, and further discusses the possible influential variables affecting WIM data quality and the gaps in current research. This

review demonstrates a description of WIM system operation and an understanding of how various factors influence a WIM system's performance. The methodologies presented for evaluating the data quality provide options to examine the WIM system performance for diverse applications, including weight enforcement programs, weight-mile tax enforcement, and bridge and pavement design.

In the data quality examining method, the departures of established property distribution could be caused by instrument failure, calibration drift, vehicle dynamics (e.g., suspension response and load configuration), station characteristics (e.g., pavement profile) and meteorology factors (e.g., wind, temperature and precipitation). System malfunction can be related to minor or major sensor health issues. Calibration drift is the actual change in the weight measurement when weight estimate is measured under the same conditions over time.

WIM data quality metrics are to monitor the weight and axle spacing accuracy of WIM systems over time. The weight accuracy includes steering axle weight, GVW distribution and accumulated ESAL. Steering axle weight is found to be independent of GVW, vehicle mix and axle spacing, and is a robust

indicator of the weight drifts. Apart from the weight accuracy metric, the axle spacing is also a key output of WIM systems. This is important in the vehicle classification algorithms and also is used to determine vehicles' compliance within the weight enforcement program. The drive tandem axle spacing is independent of the vehicle loading because of technological reasons, which qualifies it to be a data quality metric.

### **3.0 DATA DESCRIPTION**

This chapter presents the WIM data source for this research, and describes the locations of WIM stations in this research.

#### **3.1 WIM Data**

The WIM dataset is archived in the WIM SQL server at Portland Oregon Regional Transportation Archive Listing (PORTAL, see <http://portal.its.pdx.edu>). PORTAL is maintained by the Intelligent Transportation Laboratory (ITS) at Portland State University. It offers traffic data, performance measures and analytical tools in a user-friendly interface. The WIM data are provided by ODOT's Motor Carrier Division in comma-separated files. The raw files are sent to the WIM SQL server monthly, and before that they are processed and uploaded to archive. An R script was developed to retrieve the raw WIM data files from the SQL database and generate an analytical plot.

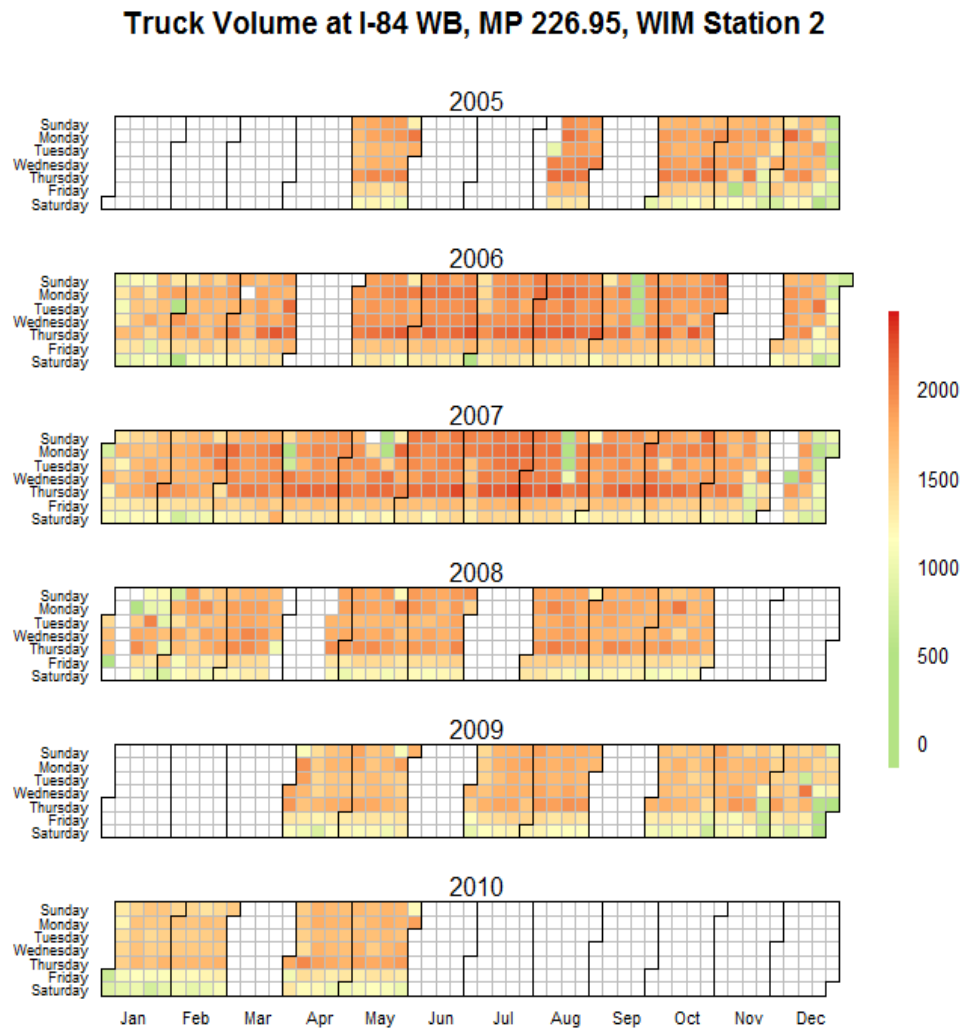
As mentioned in Section 2.2, a WIM system installed in a lane has a pair of dual-loops and a load cell sensor in between the loop sensors. The dual-loops together count the vehicle volume and for each vehicle measure the axle bypassing time. Having the bypass time and speed, axle spacing could be closely estimated. With group axle weight, the GVW estimate is automatically calculated.

Each record in the WIM database includes a timestamp, measurements of the individual axle load and spacing, speed, lane, station number and flags data as well as vehicle tag, type, length, ESAL and gross vehicle weight. Some data are measureable; others are derived, such as speed, gross vehicle weight and vehicle type. The WIM devices are also used to provide statistics on axle load spectra, Annual Average Daily Traffic (AADT) by truck class, and hourly trend of GVW by truck class, inter-arrival distribution of the traffic by truck class, and other traffic flow characteristics at the WIM station.

The WIM database in PORTAL is available from 2005-2010. The 2007 WIM data are found to be the best because it is missing the least data of the five years. The calendar heat plot in Figure 4 shows the number of records for each day in 2007 from a sample WIM station 2. The calendar heat plot is specialized in visualizing longitudinal data in a color-coded calendar. It gives an idea of the records coverage. The white grid means missing data, which could be caused by data communication or equipment problems.

The sample WIM system stations came from the six stations located on Interstate 84 in Oregon. The truck volume plots at sample stations in Figure 5 show their truck volume ranging from 0 to 3,500. Stations 3 and 4 have less traffic

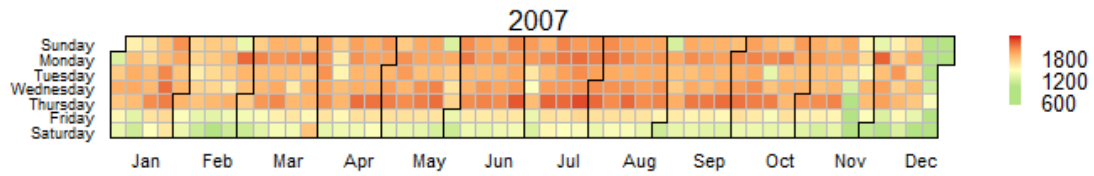
than other stations. The station locations are shown in Figure 6. With considerations of data completeness, this study selects data from January to April 2007 as sample time frame.



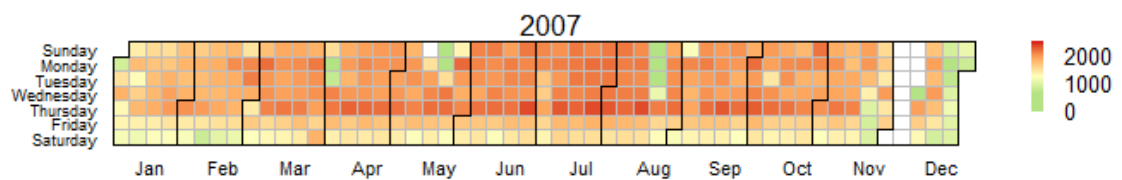
**Figure 4: Calendar Heat Map of Station 2 Truck Volume**



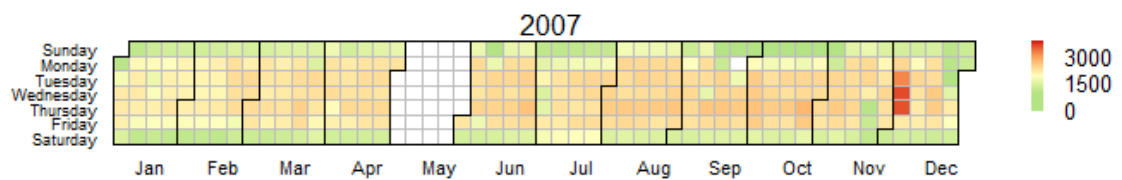
### Truck Volume at I-84 WB, MP 353.31, WIM Station 1



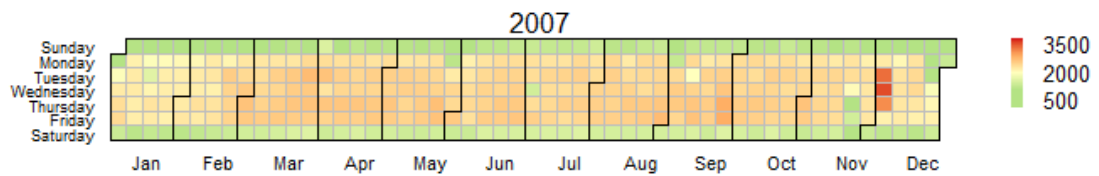
### Truck Volume at I-84 WB, MP 226.95, WIM Station 2



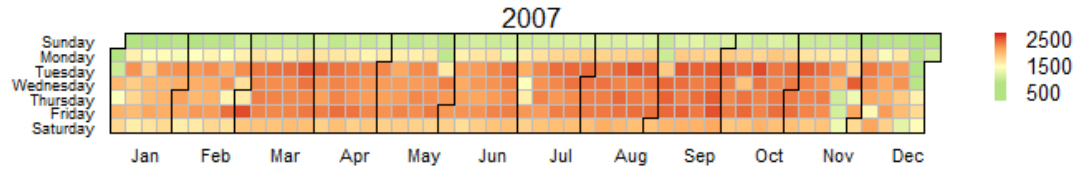
### Truck Volume at I-84 WB, MP 54.30, WIM Station 3



### Truck Volume at I-84 EB, MP 44.93, WIM Station 4



### Truck Volume at I-84 EB, MP 258.52, WIM Station 5



### Truck Volume at I-84 EB, MP 354.38, WIM Station 6

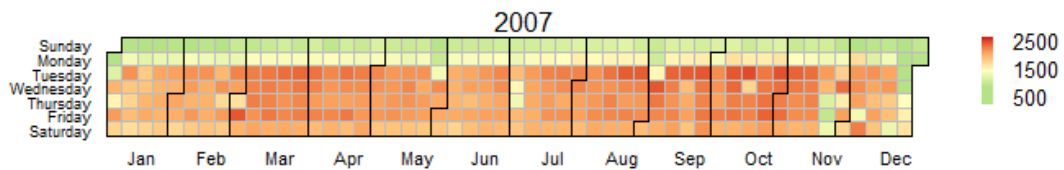


Figure 5: Truck Volume Calendar Heat Plot at WIM Stations 1, 2, 3, 4, 5 and 6

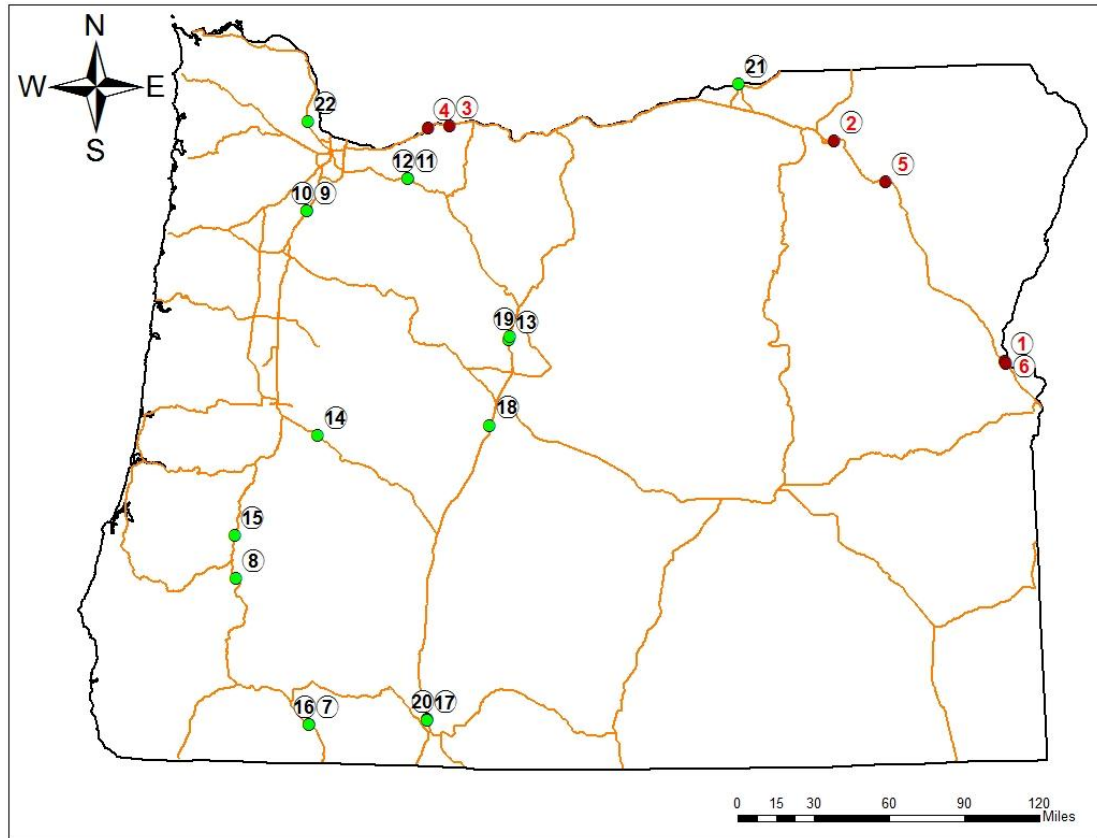
Table 1: WIM Systems Study Data Size

WIM Station	Station Name	MP	Total Records	Sample Records
			2007	Jan - Apr 2007
1	Farewell Bend POE	I-84 WB, MP 353.31	437,675	140,951
2	Emigrant Hill	I-84 WB, MP 226.95	404,413	134,998
3	Wyeth	I-84 WB, MP 54.30	351,676	126,785
4	Cascade Locks POE	I-84 EB, MP 44.93	403,634	136,618
5	LaGrande	I-84 EB, MP 258.52	462,082	155,204
6	Olds Ferry	I-84 EB, MP 354.38	458,792	155,817

The WIM data are analyzed and evaluated by week. Daily or monthly analyses may create more noise or not precisely present the data. When converting data from timestamp to week, the U.K. conversion is adopted that uses Monday as the first day of the week. Thus each selected week would have

the six-day data. The study record sizes at each WIM stations are shown as Table

1.



**Figure 6: Map of Oregon WIM Sites used in Study**

### 3.2 Static Weight Data

The static data were collected on multiple local roadways in Oregon in 2007 and 2008. WIM data could be better calibrated with a comparison of an outside source of static data with known quality. The static database includes

truck class, axle weight, observed gross weight, commodity type, trailer type, data collection time and location, but it does not record the axle spacing data. Since this data collection survey did not implement a truck transponder, the truck tags were not available to link these static weight data with the WIM data. Hence the static data are only used for comparing the weight data accuracy.

### **3.3 Summary**

This research employs an extensive amount of data from the WIM database in PORTAL. This section describes the WIM data selection criteria and the WIM system's study location, presents a WIM data consistency study, and finally introduces the static weight data.

## 4.0 METHODOLOGY

The next chapter discusses the methodologies for the performing data quality study and describes the case study. It includes a description of the procedures of four weight data quality methods and four axle spacing quality methods using the five-axle tractor-semitrailer truck fleet. The WIM data accuracy metrics in this study are a performance-based specification for WIM systems, which is more feasible to enforce than the current specification of comparing static vehicle weight with dynamic vehicle weight.

Quality checks are broadly recognized as format, chronological checks, arithmetic checks, range checks, data accuracy, and consistency across data sources. In this case, format, chronological and arithmetic checks were accomplished by ODOT's self-developed WIM data quality program. It specifically detects the errors in generic site description, file and data format, date and time sequence. Therefore, this study would mainly focus on range, consistency, and data accuracy checks.

There are several assumptions for this study. First, it is assumed that differences in site characteristics (e.g., pavement profile) among chosen WIM sites do not significantly affect the axle weight measurements. Second, the

influence of vehicle dynamics resulting from differences in suspension characteristics, tire width and pressure, and frame length in the five axle semi-trailer truck fleet is negligible. Third, it is assumed that the mix of empty and laden truck distribution do not vary at the chosen WIM stations, and the distribution can be monitored for calibration drift.

#### **4.1 Triangulation Method for Sensor Performance Measurement**

Triangulation method in this research makes uses of data from diverse WIM system sources. It is used to virtually monitor and evaluate axle weight and axle spacing data quality. As mentioned in Section 2.0, the data quality metrics - steering axle weight and drive tandem axle spacing - are independent from truck weight for the reason of mechanical design. A truck passing through multiple WIM stations is expected to have a constant value of steering axle weight and drive tandem axle spacing with little variation. To identify a truck's travel is to link the truck's RFID tag in order to trace its travel among the adjacent WIM stations. Once identified, steering axle weight and drive tandem axle spacing for each truck are obtained. In this method, only through trucks on the trip are selected. The reason is that there are usually hundreds of miles between WIM stations, so a freight vehicle has more chances to stop, rest, re-fuel or make

deliveries before it is tracked at the next WIM station. The results may change the gross vehicle weight. To eliminate this type of bias, only through trucks were selected for this study.

A truck's data quality metrics at multiple stations are regarded as a trip group. Trip data are filtered to identify the trip made by one truck passing by three ordered WIM stations, collect all trips of this type and extract through trips. The trip in this method is defined as one direction of truck travel on I-84 in a day. The route sequences can be either westbound or eastbound, which means a truck going by way of WIM stations 1-2-3 or WIM stations 4-5-6.

The ultimate purpose of this method is to compare one truck's weight data at three stations in a trip, and the station with off-weight data is the suggested malfunctioning station among the three stations. It is noteworthy that a truck's weight data at three stations must be compared in a similar condition. Therefore, after data filtering for the time restriction and highway route of each truck, it is good to analyze data in its traveling route sequence and speed. As an example, below are the data processing procedures for a westbound trip (going by way of station 1-2-3):

1. Extract the variables of timestamp, target station numbers, truck tag and data quality measurements (steering axle weight and drive tandem axle spacing) of each 3-S2 truck fleet record from the WIM database from January to April 2007.
2. Analyze each truck's travel time and route to identify the type of trucks that finished the designed trip(s) of passing station 1-2-3 in one day. For the logic coding in R, first order data by time and split data by truck tag; secondly, within each subgroup, select the records with the length of unique station numbers equal to 3, and then select the records that indicate trucks passing the three unique stations within one day; and thirdly, split each truck tag subgroup by day and select the records for those traveling by station 1-2-3 in order.
3. Select only through trips in which a truck traveled without stopping. In theory, data quality metrics are stable regardless of the changing loading weights or the changing kingpin location (caused by attaching different two-axle trailers). To further improve study accuracy, only through trips are selected for analysis. The through



trucks do not gain or lose weight by loading or unloading, which would provide a good environment for analyzing data quality metrics - steering axle weight and drive tandem axle spacing. Through trucks are reflected by a reasonable space-mean speed calculated as trip distance over truck travel time between adjacent stations (e.g., speed from station 1 to 2 and speed from station 2 to 3). On I-84, the truck posted speed is about 55 mph, and thereby trucks traveling faster than 50 mph are assumed to be through trucks.

4. Identify the “isolated” station that has either the largest or smallest steering axle weight in a trip group. Weed out the trips having an equal interval steering axle weight array. The filtering process minimizes between-group variation for accurate evaluation of WIM system performance.
5. This step assumes the steering axle weights at the station 1, 2 and 3 are normally distributed. 95 percent of the area under this normal curve lies within  $1.96 \times \sigma$  distance from the population mean, indicating the weight data beyond the confidence interval are either

inherent noise or potential error signals from malfunctioning stations. This step applies the three-sigma limits to acquire the noise or signal weights at the potential malfunctioning stations in step 4. The population mean and standard deviation of steering axle weight is calculated from whole year data at stations 1, 2 and 3. This procedure is an effective way to prevent the situation when two stations are in abnormal operation, and the third station in good operational condition is rated as the malfunctioning station.

6. Repeat from step 3 for drive tandem axle spacing data.

Next compare a truck's weight data among three stations. A metric of variation ratio is developed, as shown in Eq. 5. This is a zero-centered variable to evaluate the deviance of the third station. Calculate the variance ratio to select one station that's weight is separate from the other two stations with a 5 percent variation tolerance. If the variation ratio is less than the tolerance ratio, this station is defined as the conforming stations in this group comparison. Eventually, the number and magnitude of the variation ratio would be plotted for a visual comparison.

$$\text{Variation Ratio} = \frac{\left| x_{\text{apart}, i} - \frac{\sum_{j=1}^2 x_{j,i}}{2} \right|}{\frac{\sum_{j=1}^2 x_{j,i}}{2}} \quad \text{Eq. 5}$$

where  $i$  is trip index;  $j$  is the other two close stations' index at trip  $i$ ;  $x$  is the data quality metric.

As such, the malfunction station would be finally determined by two dimensions: one is to count the daily occurrence of one station being accounted as a potential malfunction station; the other is to measure magnitude of deviation. Drive axle tandem spacing data would be analyzed in the same methodology with steering axle weight. These two variables are measured by different sensors so that data analyses are separate. It is notable that this research also explored the more than three WIM station trips (e.g., 1-2-3-4) within a day, but only five records meet the requirements. To explore more than three stations in a trip in the future, two days can be the possible time slice to select data.

## 4.2 Statistical Process Control Chart

Statistical Process Control (SPC) is widespread in industrial and manufacturing applications and originally stemmed from Shewhart's study in the 1920s. The main purpose of SPC is to give a signal when a process mean has moved away from the target; it also gives a signal when item-to-item variability

has increased. After statistics give the signal, actions will then be taken by an engineer or a technology operator. SPC reviews the data pattern and determines if a disturbance outside a normal distribution exists. Nowadays, it has a handful of applications in the transportation field. The National Highway Traffic Safety Administration deployed the SPC chart method to examine the quality of highway crash data in order to release the high-quality data for public research and evaluation applications (Pierchala & Surti, 1999). In addition to examining crash data quality, the SPC chart method can also detect highway incidents by monitoring traffic data in real time (Turochy & Smith, 2002).

Steering axle weight and drive tandem axle spacing are two measures of WIM system quality, which is called the control statistic in SPC. Let  $M$  stand for either steering axle weight or drive tandem axle spacing, then  $\mu_M$  is the mean of  $M$  and  $\sigma_M$  is the standard deviation of  $M$ . Then the central line (CL), the upper control limit (UCL) and the lower control limit (LCL) are fixed at:

$$UCL = \mu_M + k \times \sigma_M$$

$$CL = \mu_M$$

$$LCL = \mu_M - k \times \sigma_M$$

where  $k$  is the distance of the control limits from the central line and expressed in the  $\sigma_M$  units. The central limit theory forms the foundation of a control chart and

the fixing value of three for  $k$  so that the control limit is known as three-sigma limit. The three-sigma control limits provide an economical tradeoff between the risk of reacting to a false signal and the risk of not reacting to a true signal. False alarm (Type I error) means a situation when a control chart incorrectly rejects the “in-control” point; not reacting to a true signal (Type II error) is failure to reject an “out-of-control” point. A measurement value beyond three-sigma indicates that the process has either shifted or become unstable. The mean of independent random variables is normally distributed, provided a sufficiently large subgroup size.

**Table 2: Statistical Process Control Chart Types**

Data Type	Defect Definition	Subgroup Size	Chart
Attribute Data -Counted as Discrete Events	Defect Data - Number of defects, not number of defective units	Constant	c Chart Number of Defects
		Subgroup Size	
		Variable	u Chart Defects per Unit
	Defective units Data	Subgroup Size	
		Constant Subgroup Size Usually $\geq 50$	np Chart Number of Defective Units
		Variable Subgroup Size Usually $\geq 50$	p Chart Fraction of Defective Units
Variable Data - Measured on a Continuous scale		Subgroup Size=1	$\bar{X}$ and Rm Moving Range
		Subgroup Size < 10	$\bar{X}$ and R
		Subgroup Size > 10	$\bar{X}$ and s

SPC is used to monitor and control the whole/part of a process, ensuring it operates at its full potential. The type of SPC chart is dependent upon the type of data collected as well as the subgroup size, as shown by the table above.

The control chart is regarded as a sequence of hypothesis tests: current quality measure mean is equal to hypothesis nominal mean - central line. The control limits are viewed as confidence limits. Control charts work best for numeric data with normal distribution assumptions. The  $\bar{X}$  average chart and s standard deviation chart are two types of statistical control charts that apply to continuous data with a subgroup size larger than 10. The SPC average chart and standard deviation chart are especially applicable for large sample sizes of subgroups. The two types of control charts are always used together because a sample having a large subgroup usually has a poor range of statistics on a  $\bar{X}$  average chart. According to Table 2, the average and standard deviation chart are qualified to test the quality data measure – steering axle weight and drive tandem axle spacing. If the average shifts from the process target or item-to-item variation increases, the process is out of control. The average chart has to be accompanied with the standard deviation chart. First of all, SPC control charts do

not require normally distributed data in order to work and they will work with any process distribution. In order to work with any distribution, it is important to have a measure data dispersion which is, in general, expressed by standard deviation.

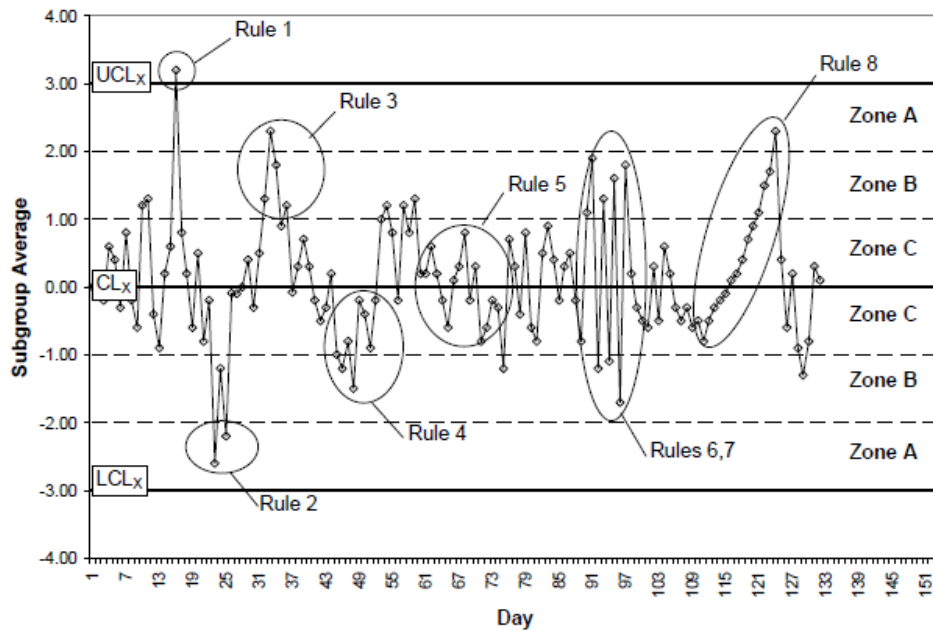
The Shewhart SPC procedures provide eight standard tests for special causes, which are also known as rules for lack of control, supplementary rules, runs tests, runs rules and Western Electric rules. These rules identify an out-of-control process, and are listed below and illustrated in Figure 7.

1. One beyond Zone A (outside of the upper control limit or lower control limit) ;
2. Two out of three consecutive points in Zone A or beyond;
3. Four out of five consecutive points in Zone B or beyond;
4. Nine points in a row in Zone C or beyond on one side of the central line;
5. Fifteen points in a row in Zone C on either or both sides of the central line;
6. Eight consecutive points on both sides of the centerline;
7. Fourteen points in a row alternating up and down; and

8. Six steadily increasing or steadily decreasing points.

Here the boundary of zone A, zone B, and zone C respectively corresponds to  $3\sigma$ ,  $2\sigma$ , and  $\sigma$ , which is calculated by the average and standard deviation of the variable. Rules 1, 2, 3, and 4 are applied to one side of the control chart; rules 5, 6, 7 and 8 are applied to both sides of the control chart. These rules improve the sensitivity of the Shewhart original three-sigma chart to small changes in a process. If the error exceeds an acceptable level, the data cannot be acted upon reliably. Most of the rules are primarily used for an average chart and a standard deviation chart. Rules 1 and 4 are the two classic Shewhart tests. For rule 4, there is no hard rule value regulating the number of consecutive points, only indicating a run of monotonically increasing or decreasing points of run length. That is, if the number of consecutive runs on one side of the control limit is beyond the predetermined run length, the part beyond the run length is violating runs. In this study, the SPC control chart method monitors quality data to test the existence of the assignable cause of calibration drift and equipment malfunction. The number of points in rule 4 will be discussed and determined in Section 4.2.2.





**Figure 7: Shewhart Rule Illustration**

*Source: Nichols, A. and Bullock, D., (2004), Quality Control Procedures for Weigh-in-Motion Data. West Lafayette, Indiana: Purdue University Purdue e-Pubs.*

#### 4.2.1 Subgroup Frequency and Sample Size

The ideal subgroup rule is to minimize the variation among the items within a sample and maximize the opportunity for detecting variation between the samples. The subgroup frequency of quality data is proposed by day for several reasons. First, if set in an hourly frequency, data collected in consecutive hours cannot reveal special cause variation. At low-volume locations, WIM sensor drift and sensor failure is slow over time; there might not be adequate traffic volume during a day, causing a gap in the control chart. Secondly, a weekly or monthly subgroup frequency has lag time in obtaining data points.

For a high level of data accuracy, it is not accurate to only evaluate data once a week or month. Therefore, the logical choice for subgroup frequency in this application is a day. It helps exhibit erratic behavior in the testing period and aid process stability. Variations due to random causes are expected to be constant from day to day. From January to April 2007, the total subgroup number is 119. Since the initial calibration for all selected WIM stations is not accessible, centerline and control limits can be calculated from the 119 subgroups.

After selecting day as the subgroup frequency, the subgroup sample size needs to be determined for the subgroup's mean and standard deviations. This requires equal sample size (i.e., daily vehicle counts) to keep constant control limits. A large subgroup sample size (100) provides a good approximate normal distribution of subgroup means. The number of subgroups can be determined assuming variation of the sample is representative of the population. Random sampling that avoids biases was implemented for this research. The random sampling that generates equal sample size does not obtain a true random sample, but sorts the 3-S2 truck records by the seconds of the record timestamp. Thus, the second that a vehicle arrives is a random occurrence.

#### 4.2.2 Average and standard deviation chart

The SPC quality control chart analysis is performed on the statistical software R. As discussed, in SPC average and standard deviation control charts are selected on the basis of quality data, which are continuous variables and have a large sample size in this case. An average chart controls the mean of a quality characteristic. This chart is accompanied by a standard deviation chart, which monitors the increase in within-subgroup variability over time. Since the “qcc” function allows a variable sample size, the upper and lower control limit would change with subgroup sample size, and the Zones A, B and C are not specified on the chart.

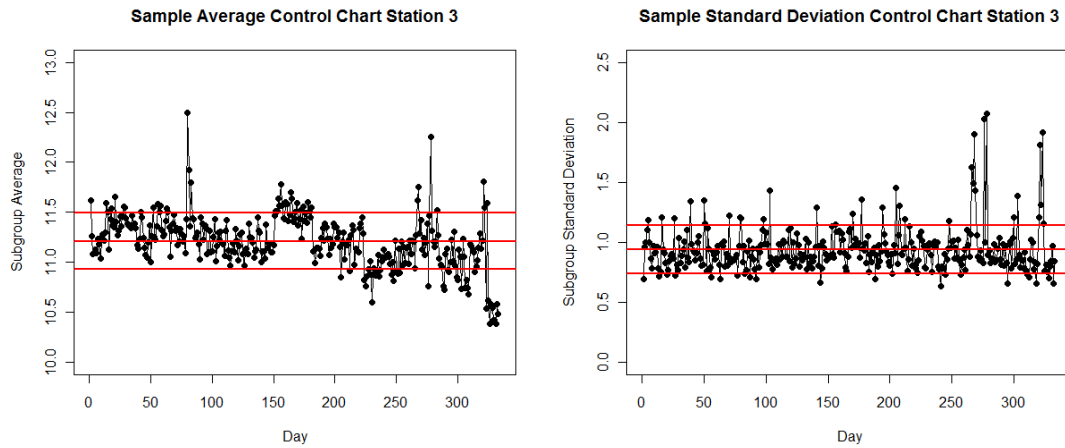
The normal distribution is the basis for the charts and requires several assumptions:

1. The quality characteristics to be monitored are adequately modeled by a normally distributed random variable;
  2. The parameters of the random variables are the same for each record in a subgroup, and each record is independent of its predecessor or successors (variation between subgroups is absent);
- and

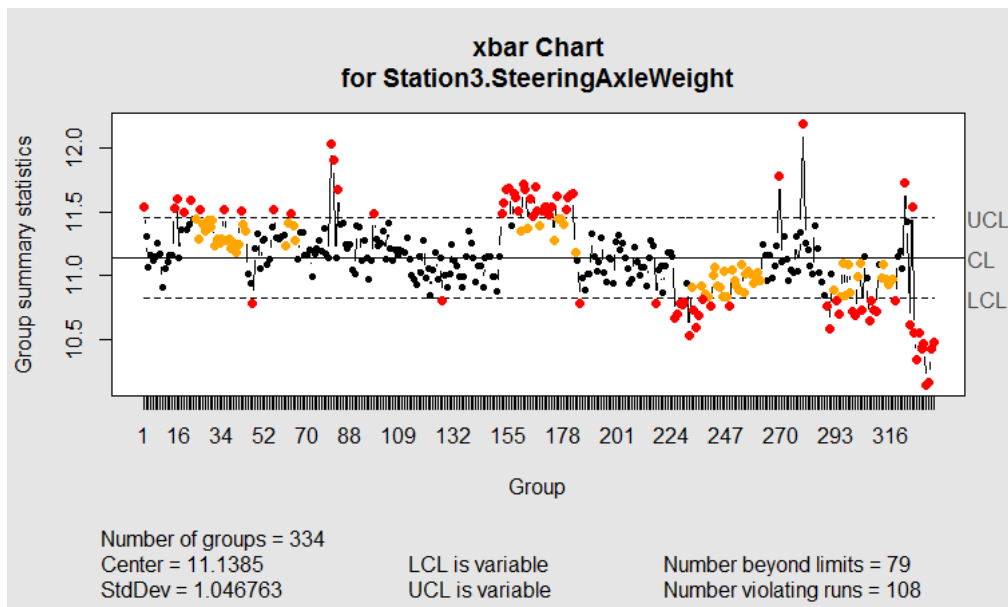
3. The inspection procedure is the sample for each subgroup and is carried out consistently from subgroup to subgroup.

The process standard deviation is using the root-mean-square estimator method, which is computed as a weighted average of subgroup estimates based on subgroup standard deviations. Control limits are established based on subgroup quality data and will not change unless a site is recalibrated. Figure 8 shows a sample average chart and standard deviation control chart with sample WIM data through real calculations of upper and lower control limit. Shewhart rules are proposed to detect the out-of-control data, but it is challenging to conduct the statistics tests on the plots. Therefore, this research adopts the SPC package `qcc` in R to plot control charts with zones and rules. The `qcc` function in the package plots the classic Shewhart rule chart and counts the number of beyond limits and violating runs on the chart. The beyond limits are these beyond  $\pm$  three-sigma control limits on the classic Shewhart chart; the violating runs follows the rule that seven out of seven are on the same side of the chart referring to the central line. The mean chart and standard deviation chart generated by R's `qcc` package don't specify the violating run index, but a function "`shewhart.rules`" in R could signal the out-of-control points highlighted

in Shewhart charts. When a point violates the Shewhart rule and stays beyond limits, it would be shown in red as illustrated in Figure 9.



**Figure 8: Sample Average and Standard Deviation Control Chart**



**Figure 9: Sample Average Control Chart using QCC Package of R**

When a point falls outside of the established limits on the control chart, an underlying process is to determine whether a special cause has occurred. When no special cause occurs in this process, a process is considered in control. However, an in-control process plotted on a control chart can still signal a special cause at a low probability. If the process is a normal distribution, 99.7 percent of the population is captured by the curve at three standard deviations from the mean; in a control chart, there is approximately a 0.3 percent probability of a point exceeding the three-sigma control limit once every 1/0.3 percent observations (370 observations). Shewhart charts are good at detecting a large change in the process mean or variance. The in-control Average Run Length (ARL) on Shewhart is set to eliminate that Type I error (incorrect reject of null hypothesis). The term ARL means a mean number of times a process will have been sampled before a shift in this process level is signaled by a control chart. Champ and Woodall (1987) provided a method of obtaining the exact ARLs for the Shewhart chart with run rules for normally distributed quality characteristics, as shown in Table 3. It gives the ARL values for an average control chart with certain run rule combinations. The higher the run length, the more acceptable the data are. When a process is operating at an unacceptable level, the probability of

accepting the process should be low and a signal in this process level should come from the control chart with a smaller number of subgroups drawn. Therefore, the ARL is set to be seven, according to Table 3. The test rule would be seven consecutive points falling on the same side of the central line.

**Table 3: Average Run Lengths for Various Shift Levels and Run Rules**

Process level shift (in $\sigma$ terms)	Rule 1	Rules 1 & 2	Rules 1 & 3	Rules 1, 2 & 6
0 or no shift	370.352	225.410	166.034	122.036
0.5	155.205	77.715	46.176	36.171
1.0	43.889	20.003	12.663	11.725
2.0	6.302	3.646	3.680	3.502

### 4.3 Data Quality Flag

The data flag process is a critical procedure to identify and remove records with formatting mistakes, spurious data and other errors. The records that do not follow the general record pattern can be identified by a set of logical data quality rules. This study first weeded out the duplicated timestamp with another record. The erroneous records are then identified by 12 data quality rules developed by Pelphrey et al (2008) as follows:

1. Record where the GVW value is equal to 0.0;
2. Record where an individual axle is greater than 50 kips;
3. Record where the speed is greater than 99 mph;

4. Record where the speed is less than 10 mph;
5. Record where the length is greater than 200 feet;
6. Record where the sum of the axle spacing lengths are greater than the length of the truck;
7. Record where the sum of the axle spacing lengths are less than seven feet;
8. Record where the first axle spacing is less than five feet;
9. Record where the number of axles is greater than 13;
10. Record where the GVW is greater than 280 kips;
11. Record where any axle spacing is less than 3.4 feet; and
12. Record which has a GVW  $\pm$  the sum of the axle weights by more than 7 percent.

WIM sensor health can be communicated by the 12 logical rules above, where rules 1, 2 10 and 12 are used to diagnose weight sensors and others for loop sensors. The data cleaning and filtering process checked and flagged these 12 data quality rules for each record. The sample dataset include the flagged data records. For each station's database, a script summed all logical data quality flags

$$weighted\ flag = \frac{\sum_{j=1}^V flag_{j,i}}{number\ of\ rules \times V_i} \quad \text{Eq. 6} \quad 69$$



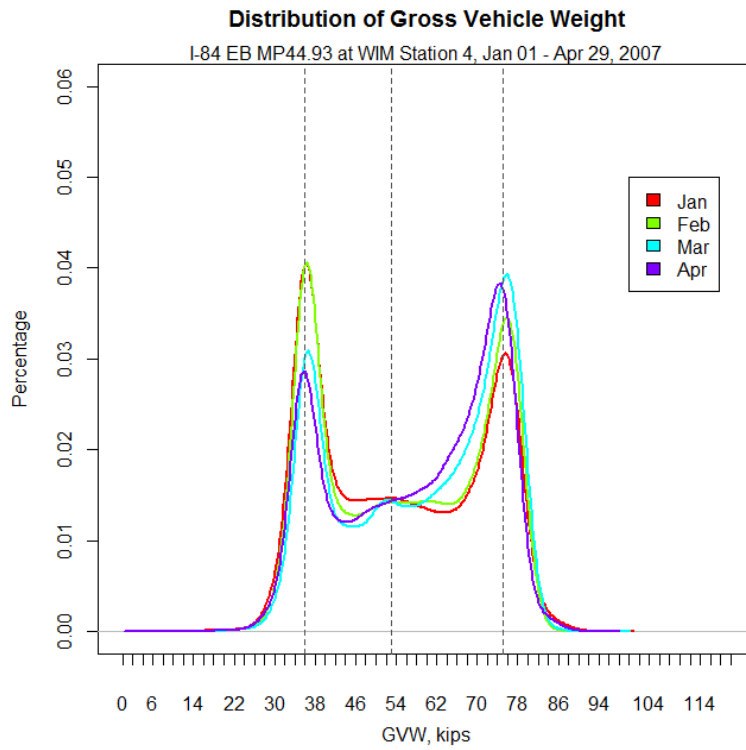
for each record as a new column and then combined the sum of flags by day. The derived daily sums of logical flags were weighted by the number of rules applied and the traffic volume at the corresponding date. The line plot would present the number of flags per day for each sample station, as shown in Eq. 6.

Where  $V$  is Class 9 traffic volume at day  $i$ , and  $j$  is the index of Class 9 vehicle: population at day  $i$ .

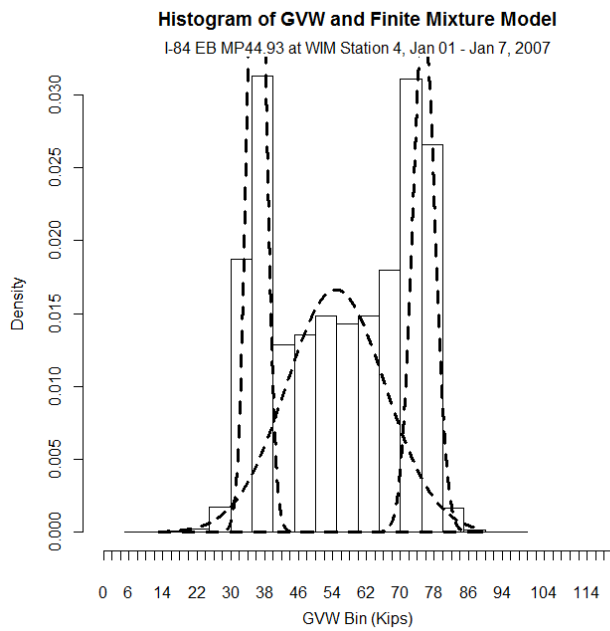
#### **4.4 Finite Mixture Distribution**

The finite mixture model is used to fit a combination of normal distributions to the GVW distribution of FHWA Class 9 trucks for WIM systems to identify peak parameters. This mixture model method is to quantify the GVW distributions of commercial vehicles and find corresponding parameters. Most WIM systems' GVW data have two or three peaks in distribution, respectively, for empty trucks, possible half-laden trucks and the full-laden trucks. The method experiments to fit GVW distribution by using bimodal, trimodal and quadmodal to find the best number of mixture components, and then calculates each component's parameters by using R's "mixtools" package. After identifying the mixture model, plot the component mean and proportion.

Take Station 4 as an example. The GWV distribution of Station 4 is shown in Figure 10. As can be seen in the plot, Station 4 is a classic GVW distribution, having three peaks at 36 kip, 53 kip and 75 kip. After acquiring the number of components in mixed distribution, the mean and proportion of each weekly subgroup can be calculated by the `normalmixEM` function in R's `mixtools` package. The `normalmixEM` function requires the number of components, initial values of mixed proportions, component means and standard deviations for an algorithm and also needs to declare the convergence criterion. After hundreds of iterations, R provides the mean and proportion of each subgroup. Figure 11 provides an overview of how the selected mixture model closely fit the empirical GVW distribution.



**Figure 10: Distribution of GVW in Station 4**



**Figure 11: Histogram of Mixture Distributions in Station 4**

#### **4.5 Axle Spacing Error Rating Model**

This method considers the variations of the drive tandem axle spacing, which can be categorized into two parts: assignable variation and chance variation. Assignable variation can be discovered and corrected at the sensor level and also is dependent on environmental effects, such as temperature, on measurement systems. The other variation measured - chance variation - is inherent in the measurement process and cannot be controlled and blocked. A significant spike is attributed to the assignable cause.

This method uses a term error rating, ER23, to quantify the loop sensor soundness, which is defined as the zero-centered interval containing 95 percent axle spacing measurement errors. This research requires an assumption that the true drive tandem axle spacing and axle spacing measurement errors are normally distributed. From a review of millions of WIM-measured drive tandem axle spacing records, when the WIM system is in good operating condition, it is concluded that the true drive tandem axle spacing is approximately 4.2 feet with a rough variation of 0.03.

The measured mean percentage  $M$  and standard deviation percentage  $S$  are values for the total error, which are used for calculating mean percentage error  $m$ , standard deviation of percentage error measured by WIM  $s$ , and standard deviation percentage of measurement error on drive tandem axle spacing,  $s'$ .

$$m = \left( \frac{M - 4.2}{4.2} \right) * 100 \quad \text{Eq. 7}$$

$$s = \left( \frac{100}{4.2} \right) * S \quad \text{Eq. 8}$$

$$s' = \sqrt{s^2 - 3^2} \quad \text{Eq. 9}$$

A large sample size ( $n > 1,000$ ) can simplify calculation of ERas23 by using the values of  $m$ ,  $s$ , and  $s'$ . For a mean error of zero,  $\delta = 1.960 * s'$ . A mathematical relationship between ERas23 is developed through multiple linear regression analysis as below. Here, if  $s < 3$ , then  $ERas23 = |m|$ ; if  $s \geq 3$ , calculate the standard deviation  $s'$  of the measurement error  $s' = \sqrt{s^2 - 3^2}$ . For  $s' < m$   $ERas23 = |m| + (1.645 * s')$ ; for  $s' \geq m$   $ERas23 = (|m| + 1.645 * s') * (0.387\alpha^3 - 0.928\alpha^2 + 0.759\alpha + 0.789)$  where  $\alpha = \tan^{-1} \frac{s'}{|m|}$

Based on the analysis of standard deviation of measurement error and mean of measurement error, Slavik and Wet (2012) found Eras23 did not exceed

the maximum acceptable error rating of seven in good operating condition. The 7 percent was selected as a threshold.

This metric merely identifies the axle spacing accuracy of Oregon's WIM system because of the completely separate measurements for steering axle weight and drive tandem axle spacing.

## **5.0 ANALYSIS AND RESULTS**

The following chapter represents the results of data quality analysis. The first section provides a descriptive statistics study of WIM data, and compares WIM weight data quality metrics with the static weight data. The second and third section, respectively, explore the weight and axle spacing data accuracy of FHWA Class 9 trucks at the selected WIM devices. The fourth section quantifies and compares the results of weight and axle spacing data accuracy. Open source software tool R is used for the entire study.

### **5.1 Descriptive Statistics**

Steering axle weight and drive tandem axle spacing are two variables independent of the loading of a truck. Steering axle weight is measured by load cells; drive tandem axle spacing is measured by two loop sensors. The weight and axle spacing data quality could be different due to the two metrics being measured from different sensors. This section uses the steering axle weight and drive tandem axle spacing to analyze the weight and axle spacing data quality.

The steering axle weight and drive tandem axle spacing are theoretically consistent for the reason of mechanical design. The sample 3-S2 truck fleet data were retrieved from WIM station 1, 2, 3, 4, 5 and 6 databases on I-84 in 2007.

Table 4 shows the average axle weight and spacing. The average coefficient of variation of steering axle weight is 9 percent, much more steady than other axle weight. In terms of the axle spacing, the average coefficient of variation for drive tandem axle spacing is 8.8 percent, even higher than the axle spacing between the steering axle and first drive tandem axle. The reason that drive tandem axle spacing varies more is due to the kingpin (the point at which the semi-trailer connects to the tractor), as shown in Figure 12. The drive tandem axle spacing of all FHWA Class 9 vehicles is constantly 4.2 feet with an 8.8 percent variation, but the location of the kingpin is changeable. The trailer length is restricted up to 53 feet. When setting the kingpin close to the cab, air resistance can be reduced in favor of fuel savings. The spacing between the steering axle and the middle of the drive tandem axles ranges from a cabover long-haul tractor of 12.2 feet to a conventional long-haul tractor of 19.5 feet. Once the kingpin is fixed, its drive tandem axle spacing should be constant with a little variation.

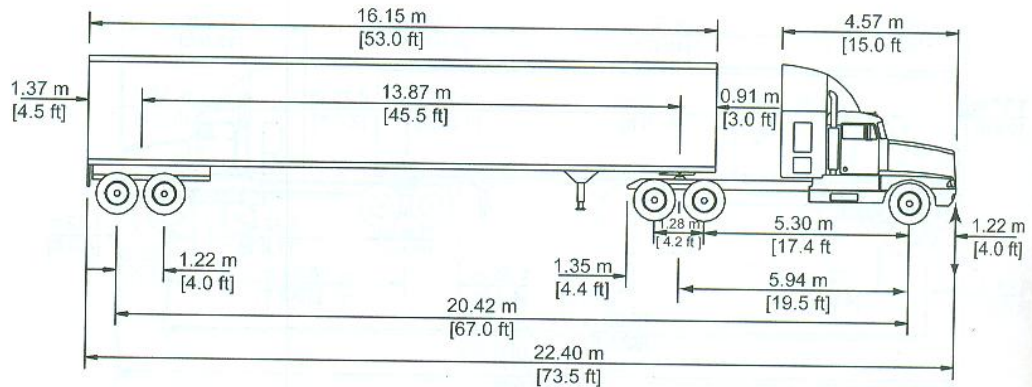
**Table 4: Coefficient of Variation of Average Axle Weight and Spacing at WIM Stations**

	Axl1	Axl2	Axl3	Axl4	Axl5
CV	0.091	0.264	0.277	0.356	0.407
	Spc1	Spc2	Spc3	Spc4	Spc5
CV	0.085	0.088	0.090	0.110	NA

**Table 5: Coefficient of Variation of Average Steering Axle Weight at Static Data Collection Sites**

	Myrtle Point	Juniper Butte	Wilderville	Burns Junction	Rocky Point	Cold Springs
2007	0.106	0.065	0.082	0.065	0.103	0.080
2008	0.085	0.055	0.094	0.063	0.084	0.073





**Figure 12: Length of Commonly Used Semi-Trailer Truck**

Many states adopt the traditional WIM weight accuracy approach with a pre-weighted truck fleet. Instrument trucks with a known static weight make multiple passes over weigh devices to determine calibration factors. These calculated factors adjust WIM outputs to be approximately close to static weights. This traditional approach, however, is not practical: the static scale has to be adjacent to WIM sites; the dedicated test trucks are not representative of the mixed normal traffic at a WIM site; its operation is costly to perform in terms of fleet scheduling and crew cost on regular basis. The analysis employs the static data source to evaluate the accuracy of the overall weight data of WIM systems.

As a reference, the static load data of a 3-S2 truck fleet was explored, shown in Table 5. Among 19 static data collection sites, the closest sites to I-84 are Rocky Points and Olds Ferry. However, Olds Ferry was removed because the

data collection was conducted only on Saturdays, which would significantly bias the commercial vehicle data. The coefficient of variation of static steering axle weight at Rocky Points was 10.3 percent, but the coefficient of variation of the static steering axle loads at six sites within a fleet of 3-S2 were small, around 7.8 percent. The static weight variation data is around 9.1 percent, the weight variation of a WIM system.

#### **5.1.1 Truck Characteristics**

This section selects five truck characteristics to analyze data tendencies and the relationships of these tendencies. The data contains steering axle weight, GVW, drive tandem axle spacing, truck length, and speed. Weight data and spacing data are measured by load cell sensors and inductive loop sensors. Thus, the steering axle weight data and GVW data are expected to be closely associated, as are the drive tandem axle spacing data, truck length data and speed data.

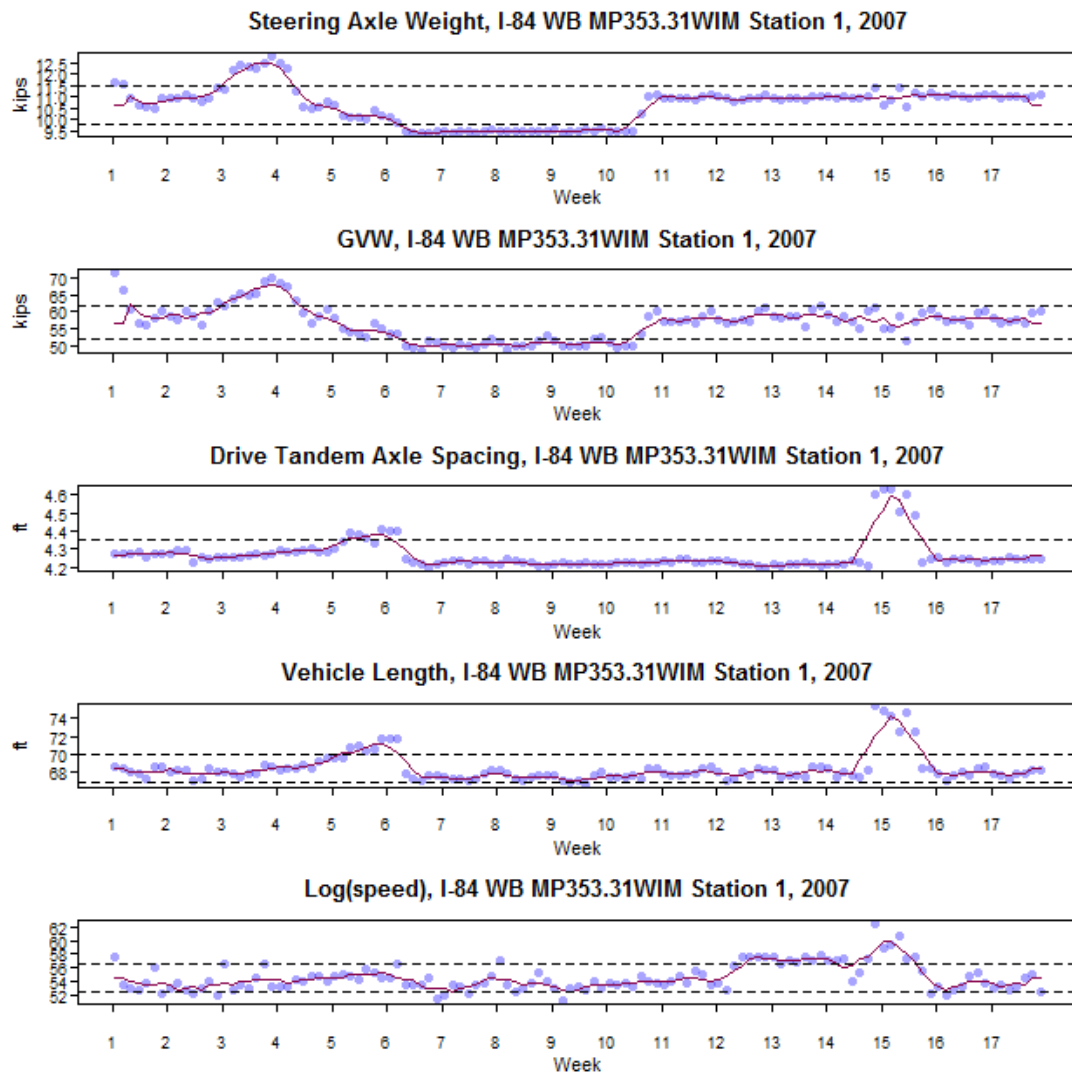
The arithmetic moving average plot for the five variables is shown as a red line in Figure 13 to smooth the data in blue dots and account for the high level of data variation. The control limits are at  $\pm$ three-sigma standard deviation. The speed distribution graph uses a logarithmic y axle to present the obvious

departures from the speed distribution: a unit increase in  $x$  axles corresponds to  $\log_{10} \text{speed}$  increase in  $y$ , which exponentially decrease the  $y$  change and zoom in small fluctuates. A WIM sensor in failure-free operation and non-congested traffic is expected to have uniformly distributed steering axle weight and drive axle spacing data. A deviation from established distribution suggests erroneous data caused by a system malfunction or incorrect calibration factor.

This plot shows that weight data correspond to each other precisely; the drive tandem axle spacing distribution and truck length distribution correspond to each other closely; and speed data have a similar trend with drive tandem axle spacing and truck length, but does not match closely. As mentioned in Section 2.0, the calculated speed is used to determine the axle spacing for that vehicle. The calculated speed and the time between consecutive axle spacing crossing the first inductive loops are used to calculate axle spacing. Time measurement by inductive loops, however, is not accurate since loop detectors do not turn on and off crisply as Class II piezoelectric sensors. This might explain why speed spectra are not close to truck length and drive tandem axle spacing spectra. The drive tandem axle spacing has been used for checking speed data calibration. Among the 20 Oregon WIM stations, four stations had the reverse performance in 2007:

spikes on the speed spectra were greater than the positive fluctuations of drive tandem axle spacing and truck length, particularly in Stations 9, 14, 16, and 18. The moving average plot for WIM Station 1 to Station 6 is in Appendix A. These graphs suggest, in essence, speed spectra.

In summary, for the weight data the moving averages of GVW have more noise than the steering axle weight; for the spacing data, drive tandem axle spacing is most robust. Therefore, this study selects the steering axle weight and drive tandem axle spacing as measurement for data quality analysis.



**Figure 13: Time Series Moving Average Plot of Class 9 Trucks at Station 1**

## 5.2 Weight Data Accuracy Results

This section discusses the analysis and results of WIM weight data accuracy. A week-quantified weight data quality is explored in the method of

triangulation, SPC, data quality flag and finite mixture of distribution. A final comparison of weight data quality methods is discussed in Section 5.4.

### **5.2.1 Triangulation Method**

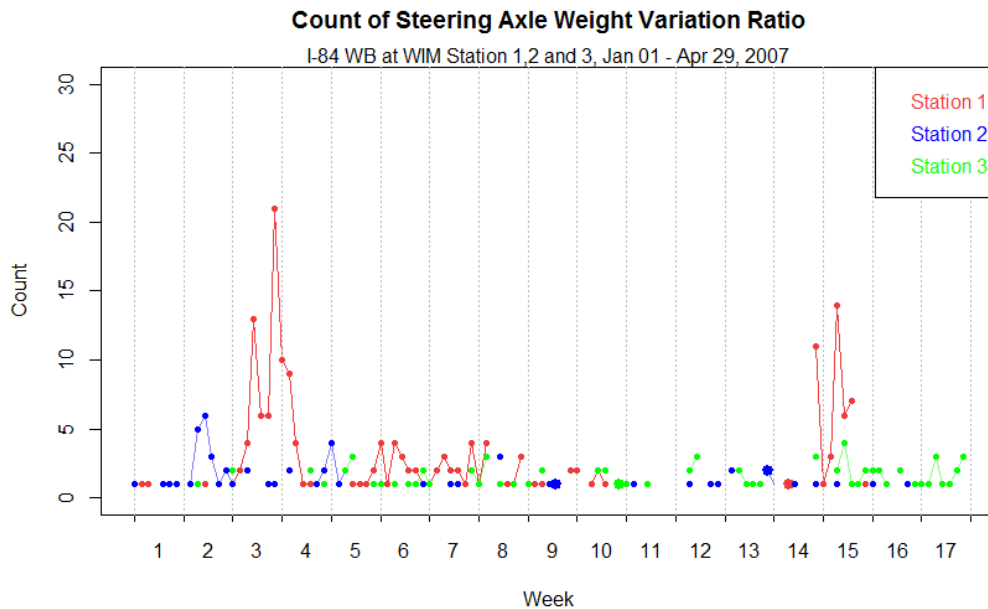
This method compares a truck's steering axle weight at three WIM stations. A station that has steering axle weight reading apart from the other two is considered as the potential malfunction station where the variation ratio is the measure of the steering axle weight deviance at the malfunctioning stations. As mentioned in Section 4.1, this method assumes the theoretical steering axle weight of a truck is approximately equal to the average of the other two close steering axle weight data.

Figure 14 shows data quality analysis using steering axle weight at Stations 1, 2 and 3 from January 1 to April 29, 2007. It clearly represents Station 1 has largest variation ratio counts from week 3 to 8 and week 14 to 15, and Station 3 has the greatest variation ratio counts at week 17. To compare the deviance in the other dimension – variation magnitude - Station 2 has a large variation value at week 9 and week 13; Stations 1 and 3, respectively, have one at weeks 14 and 10.

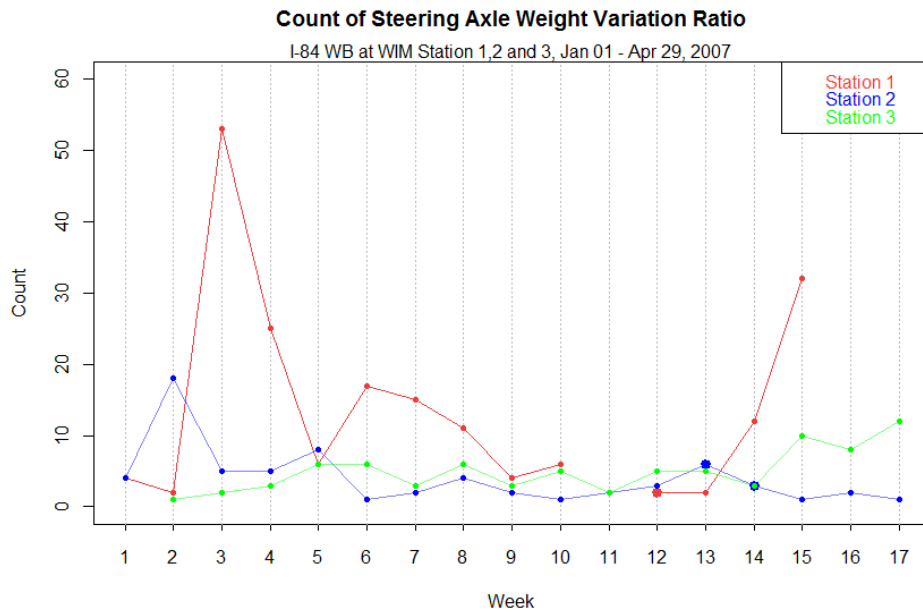
Figure 15 shows data quality analysis using steering axle weight at Stations 4, 5 and 6 from January to April, 2007. It indicated that no station has the

absolute best weight sensor performance, considering both variation count and magnitude. Station 4 has the largest variation ratio count at weeks 2 and 13. During the 17 weeks, Station 5 has the worst performance in terms of variation ratio counts and magnitude. Station 6 performed worst at weeks 6 and 9.

To facilitate the method comparison, the daily triangulation method graphs are later plotted in week level and employ the summation of weekly variation ratio counts, as shown in Figure 14 (b).

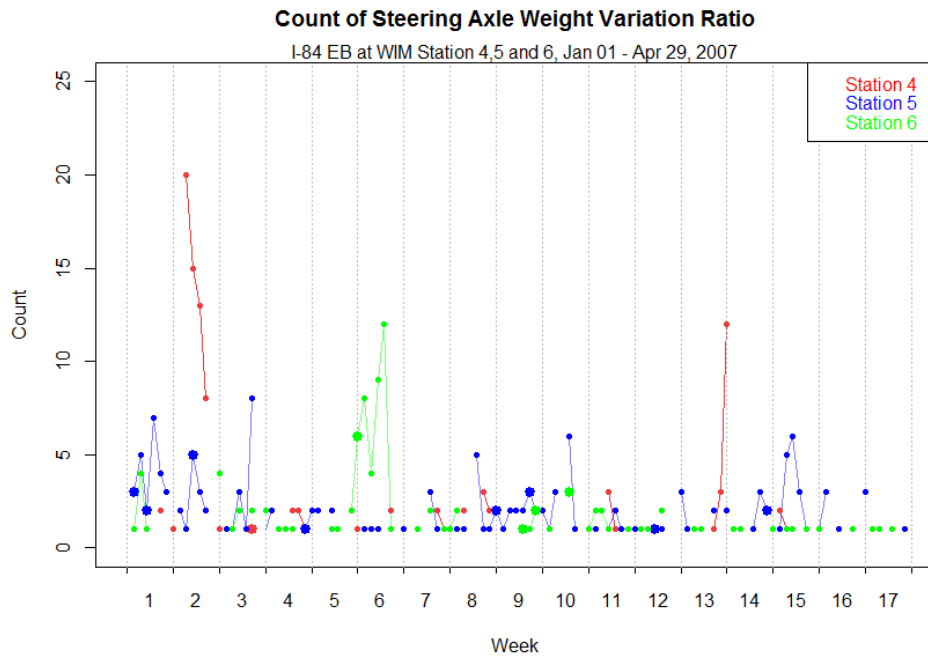


(a)



(b)

**Figure 14: Steering Axle Weight Analysis in Triangulation Method at Stations 1, 2, and 3**



**Figure 15: Steering Axle Weight Analysis in Triangulation Method at Stations 4, 5 and 6**



### 5.2.2 Statistical Process Control

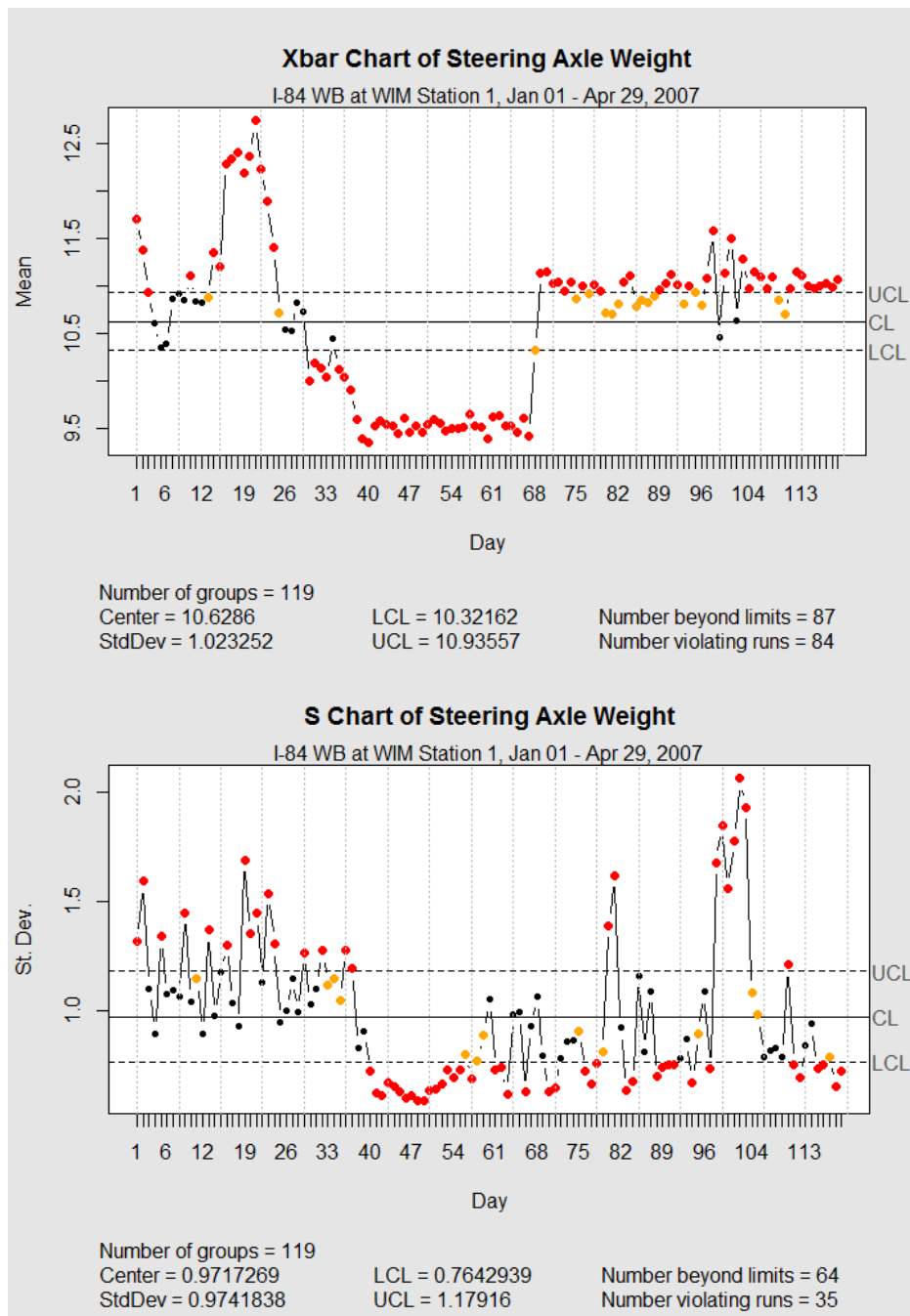
This section implements an average control and standard deviation control chart of the SPC method to plot weight quality metric data over time and identify the assignable cause variations. The mean  $\bar{X}$  chart and standard deviation  $s$  chart monitor the process mean and process standard deviation of a data quality metric in a continuous measurement scale. They are often paired in use and allow the relatively large sample size. The large sample size selected in this study contributes to detecting the true drifts in the data because the effects of outliers are reduced and the sample mean is a closer estimate of the true population mean. This section takes Stations 1 and 3 as representative to illustrate the analysis and results of SPC.

Figure 16 shows an average and standard deviation control chart of steering axle weight data at Station 1 and the Shewhart rule violations. On the average control chart, there were a total of 84 rule triggers and 87 points staying outside of the control limit. The statistics at the bottom of the chart have a fixed value for upper and lower control limit, indicating all the daily subgroups are not missing. Otherwise, the control limit would be shown as “variable,” as shown in the control chart plot of Station 2 in Appendix B. Since more than seven

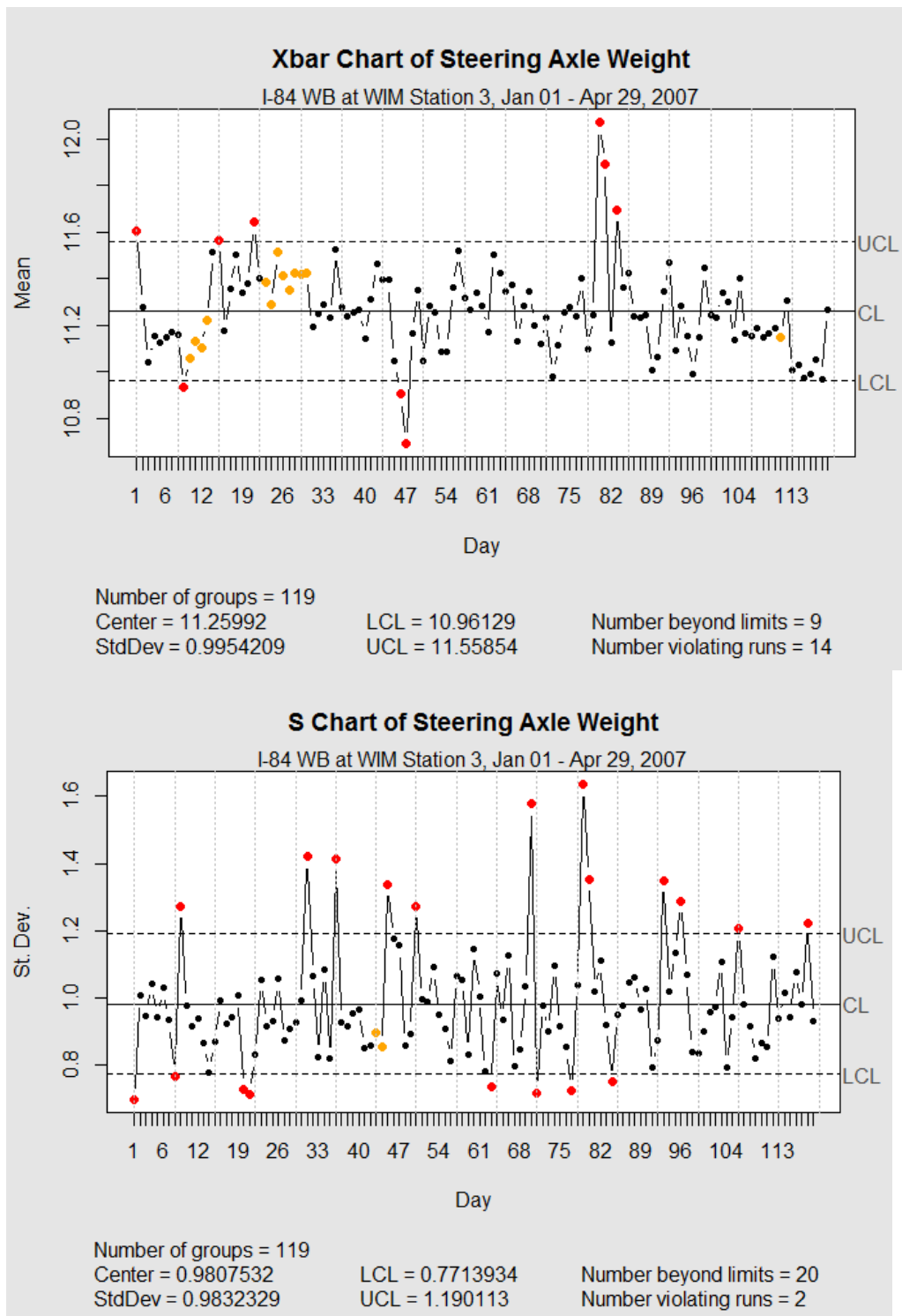
consecutive points in a row on one side of the central line is identified as an out-of-control process, the first violating run started from day 13 and ended on day 25. The points between days 14 to 24 were both beyond limits points and violating runs. The next consecutive subgroup plots beyond one side of the central line were from days 34 to 68. The rule was violated again from days 75 to 98, and it was not reset until days 106 to 119. On the standard deviation chart, there were a total of 35 rule triggers and 64 beyond-limits subgroups. It shows from days 40 to 56, there was a section of standard deviations closing to zero, which matches with the subgroup averages that shifted down on the weight control chart. It suggests the process average may actually be close to 9.5 feet at the beginning of the sixth week and the end of the eighth week.

Figure 17 shows an average and standard deviation control chart of steering axle weight data at Station 3. The average control chart represents two slow, short-term, process mean shifts at the second and fourth weeks, respectively. A violation point at week 4 is close to the central line. It suggests the true process mean may be slightly higher than the plotted process mean. If the process control line shifts up slightly, some Shewhart rule violations would be diagnosed as normal records. The standard deviation chart indicates 20 points

beyond the control limits, distributed averagely above the upper control limit and below the lower control limit. This is expected based on the inherent data variability.



**Figure 16: Steering Axle Weight Analysis in Average and Standard Deviation Chart of SPC Method at Station 1**



**Figure 17: Steering Axle Weight Analysis in Average and Standard Deviation Chart of SPC Method at Station 3**

### 5.2.3 Data Quality Flag

The cumulative load sensor flags are weighted by the number of weight flag type and the related day's traffic volume. The data quality performance of weight flags at Stations 1, 2 and 3 are as shown in Figure 18, and the weight sensor flag plot for Stations 4, 5 and 6 are in Figure 19.

At Station 1, four days in the 15<sup>th</sup> week have weight sensor flags, which may be attributed to sensor error or drift. Station 2 has one weight sensor flag in the middle of the eighth week. Stations 3, 4 and 5 have no weight sensor flag. At Station 6, the first day of the sixth week has a weight flag.

These results suggest that the test logics on load sensor are not effective to evaluate the quality of weight data. The four types of weight sensor health checks would only examine a part of the weight data quality. From two previous data analyses, Station 3's weight sensor performed well, which supports the station's evaluation results in this analysis. For Stations 4 and 5, however, there are some problems throughout these selected 17 weeks which are not detected by this analysis method.

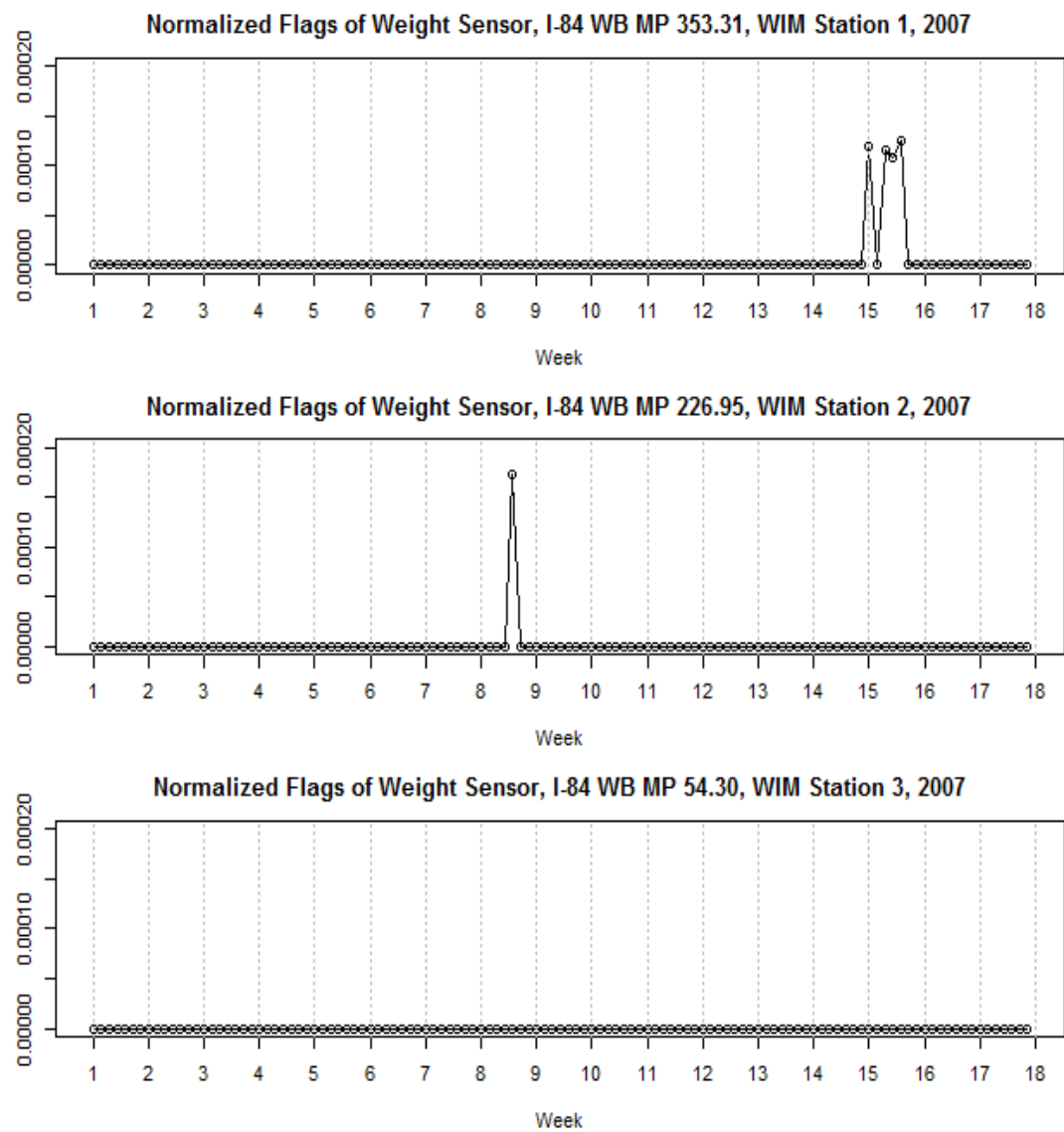


Figure 18: Normalized WIM Weight Sensor Flags at Stations 1, 2 and 3

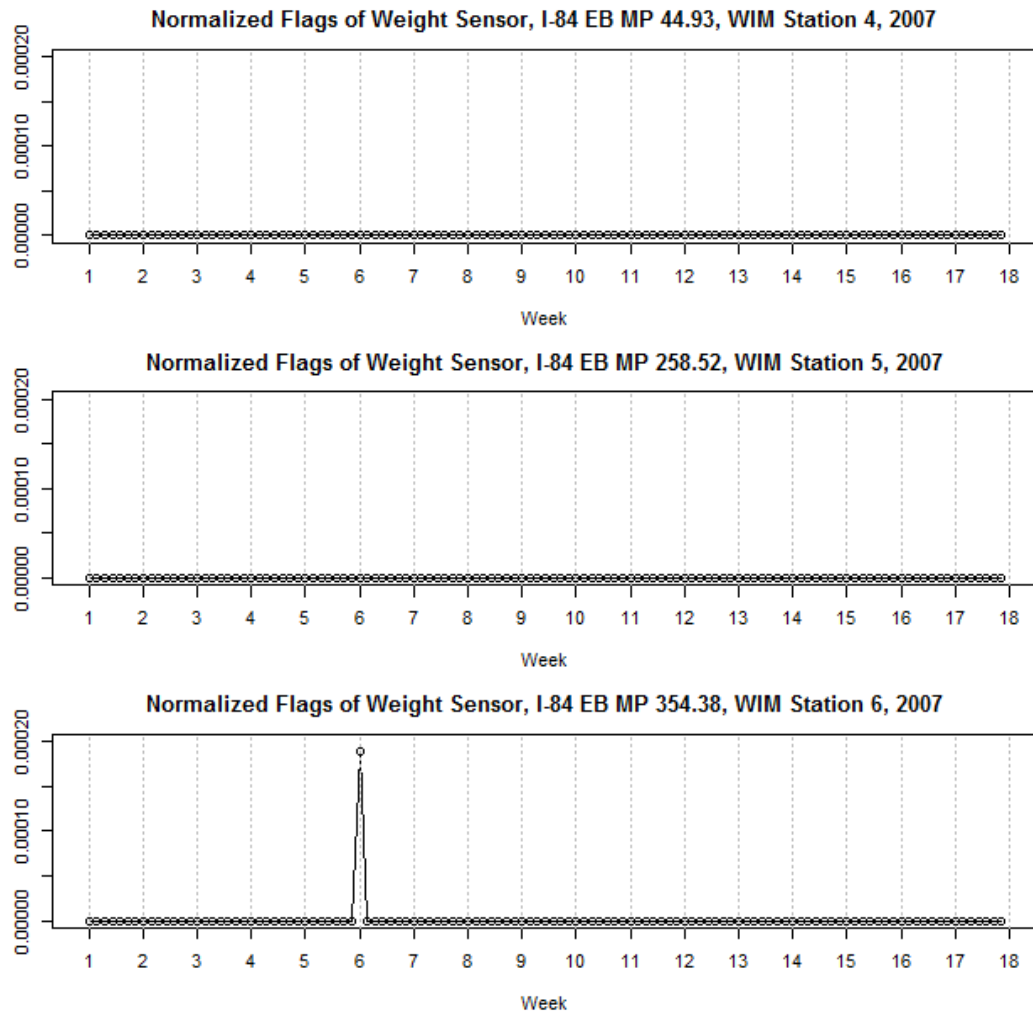


Figure 19: Normalized WIM Weight Sensor Flags at Stations 4, 5, and 6



## 5.2.4 Finite Mixture Distribution

The trimodal mixture distribution study of GVW normal distribution is conducted. This research analyzes the tendency of the mean and proportion distributions of the fitting mixture distribution.

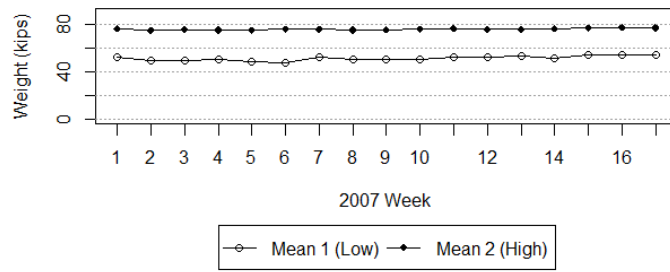
Before extracting the peak mean and proportion values from the mixture distribution sample stations, this research particularly tested the optimal number of mixture distribution for each station. Most GVW distribution is trimodal, representing for empty, half-laden and full-laden trucks. This analysis tests the sample stations in bimodal, trimodal and quadmodal. WIM Station 5 is an example of different mixed components tests, as shown in Figure 20. The graph indicates Station 5 does not tend to be better predicted in bimodal or quadmodal. To identify the best number of components, an Akaike's Information Criterion or Schwarz's Bayesian criterion can be the best option, but they are not necessary to conduct in this analysis because a reasonable number of components can be selected based on the plot of component mean and proportion.

The calibration drift is expected to be shown as a simultaneous shift occurs in three GVW component distributions, and the vehicle proportions in the low and high areas are expected not to be affected. This case is not represented in

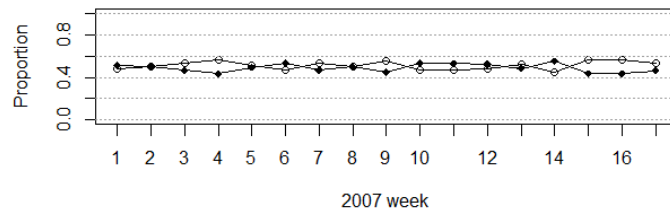
the six WIM sample stations, meaning that from January to April 2007 these six WIM systems have not been calibrated.

Station 2's distribution of the proportion changes and mean shift are shown in Figure 21. It illustrates that the low and mid peaks were gone, but the high mean was relatively unchanged. The unloaded and half-laden vehicles have lighter axles than do loaded vehicles, and not all of their axles were being well detected. As a result the entire steering weight measurements shifted up, and the proportions of high mean and mid mean dropped. The other mixture model plots are shown in Appendix C.

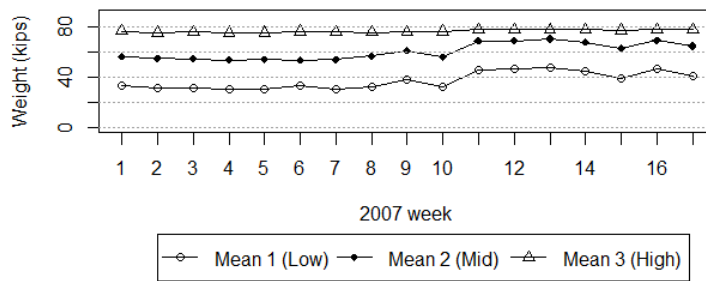
**Means of Two Components at Station 6**



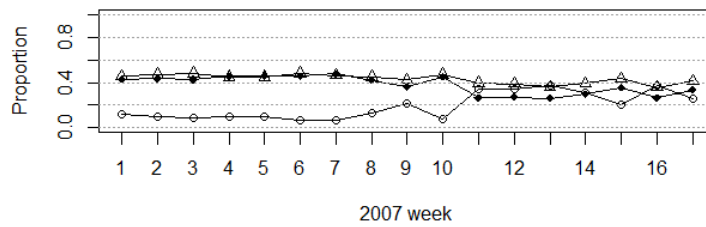
**Proportions of Two Components at Station 6**

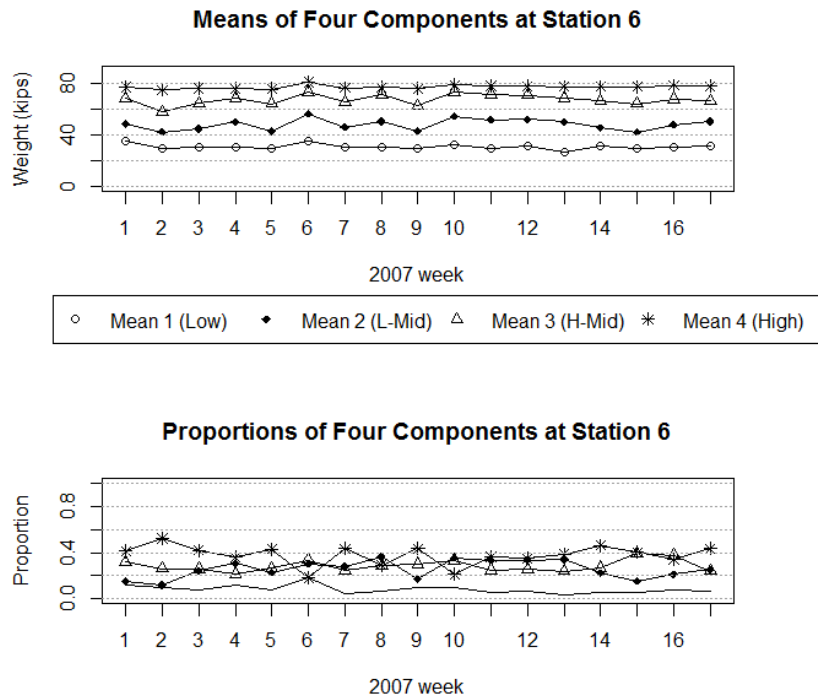


**Means of Three Components at Station 6**

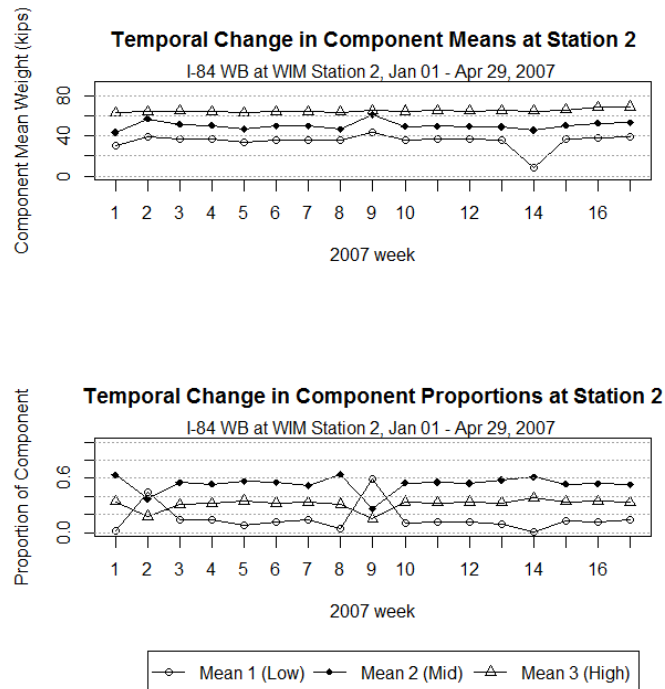


**Proportions of Three Components at Station 6**





**Figure 20: Two, Three and Four Components in Mixture Model in Component Mean and Proportion Distribution**



**Figure 21: Component Means and Proportions at WIM Station 2**

### **5.3 Axle Spacing Data Accuracy Results**

Drive tandem axle spacing data is calculated by speed and timestamp data from two inductive loop sensors. That is different from the direct measurement of axle weight by done load cells. The drive tandem axle spacing data quality analysis methods are different from steering axle weight's evaluation method. This section applies drive tandem axle spacing data into the triangulation method, statistics control chart method, data quality flag and axle spacing error rating method. A final comparison of axle spacing data quality methods is discussed in Section 5.4.2.

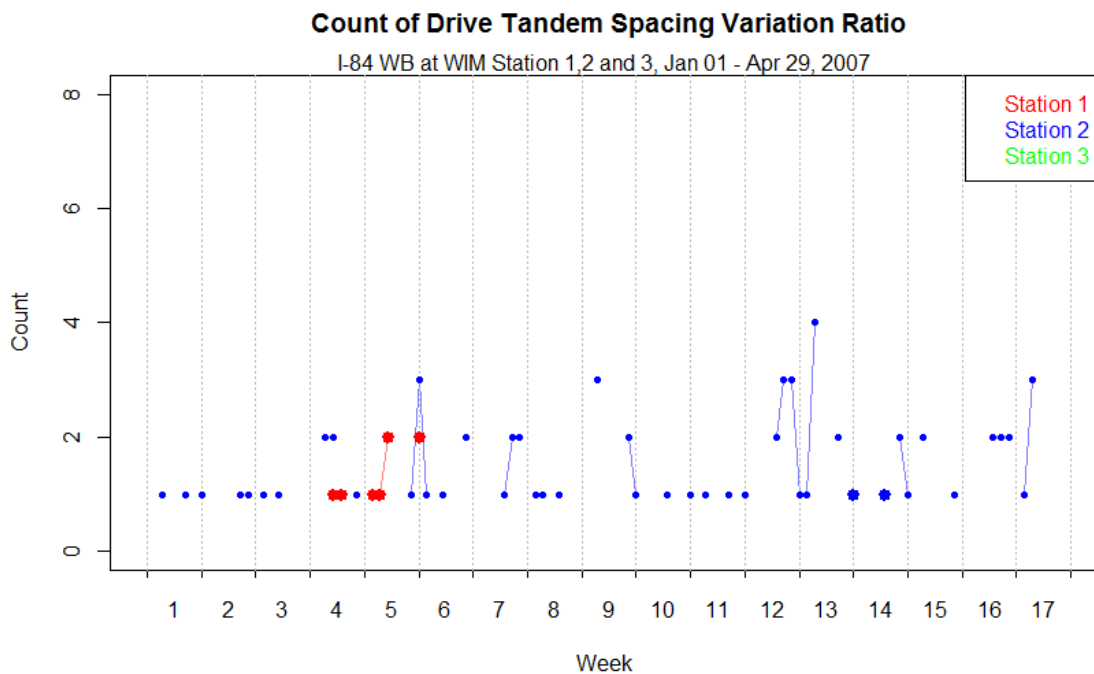
#### **5.3.1 Triangulation Method**

This axle spacing analysis process is similar with the steering axle weight analysis method in Section 5.2.1. One truck's three-drive tandem axle spacing readings from the pre-ordered stations is used to determine the variation ratio of potential malfunctioning station.

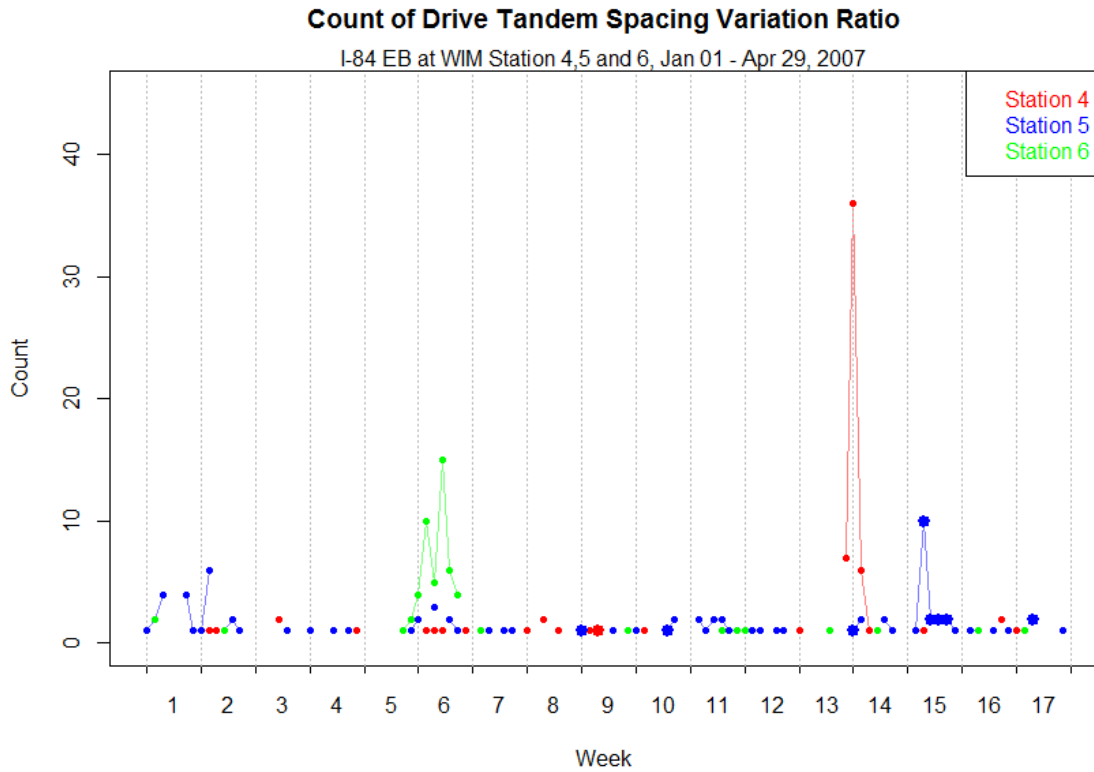
Figure 22 shows the drive tandem axle spacing data quality in Stations 1, 2 and 3. It represents that Station 2 has overwhelmingly high variation ratio counts throughout the 17 weeks in 2007. At weeks 4 and 5, Station 1 has the largest

variation ratio value among the three stations. Station 3 does not have any variation ratio.

The drive tandem axle spacing data quality analysis in triangulation method at Stations 4, 5 and 6 is shown in Figure 23. In the plot, Station 5 apparently has the most variation ratio counts. For the other two stations, Station 6 has the greatest variation ratio counts at week 5, and at week 14 Station 5's variation ratio counts has dramatically increased.



**Figure 22: Drive Tandem Axle Spacing Analysis in Triangulation Method at Stations 1, 2 and 3**



**Figure 23: Drive Tandem Axle Spacing Analysis in Triangulation Method at Stations 4, 5 and 6**

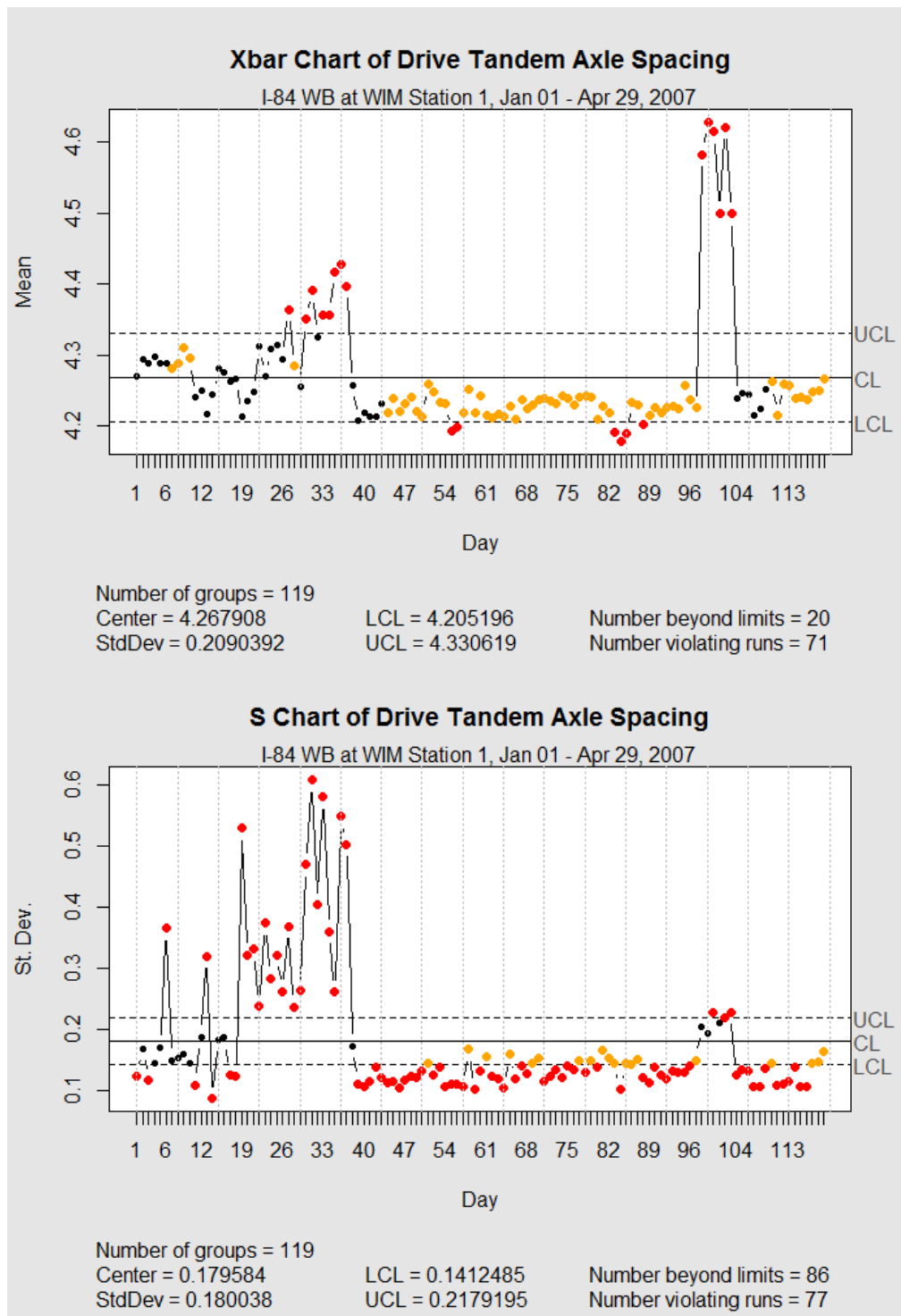
### 5.3.2 Statistical Process Control

The Shewhart alarm rule detects the out-of-control data or non-random variation conditions. Figure 24 shows the average and standard deviation chart at Station 1, where the yellow dots indicate violating run points and the red dots indicate beyond-control limits points. There were a total of 20 beyond limits and 71 violating runs. That average data in Station 1's drive tandem axle spacing data appear to be in better control than Station 1's steering axle weight data with less

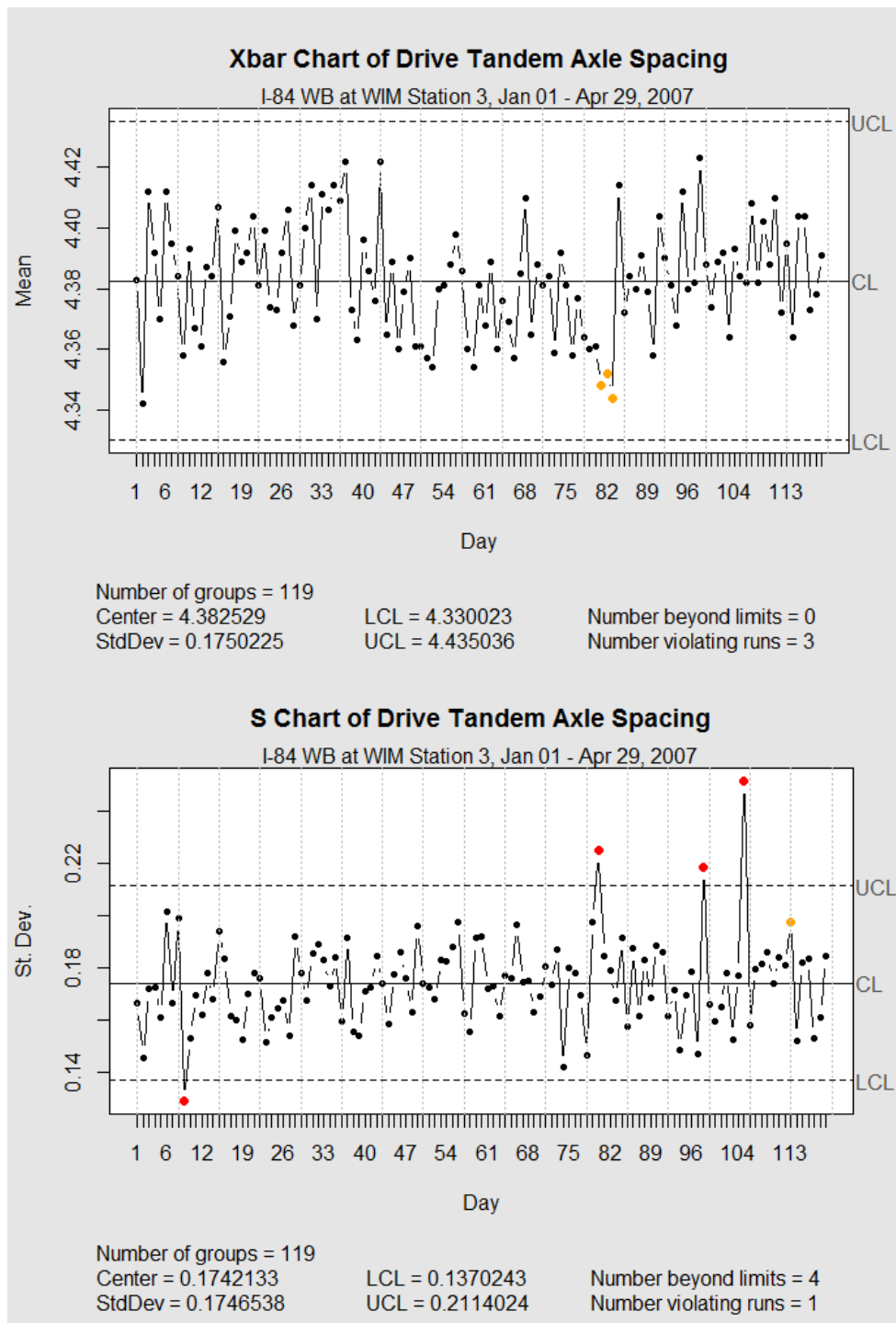
rule violations. The two charts together correspond and support each other well. The first violation run began at the seventh day of the first week and ended on the 10<sup>th</sup> day of the second week. After the next series of violating runs, the average begins to center on the lower control limit from the sixth week, which indicates the process average was shifted down to the lower control limit. If all trucks in the sample had an axle spacing of 4.2 feet and 4.34 feet, the average would be 4.9 feet exactly, and then the points closing zero standard deviation line. However, since the process is drifting down and the mean and standard deviations are close to the assumed standard deviation of 0.18 feet on which the control limits are based, the true process mean is lower than the calculated 4.34 feet mean. Hence, there is a need to adjust calibration.

Figure 25 shows the average and standard deviation chart at Station 3. The average drive tandem axle spacing data in Station 3 are in better control than Station 1 with less variation. Four standard deviations were outside of the control limits, but this is expected based on the base of common cause variation – inherent data variability. On an average control chart there are a total of three violating runs and zero beyond limits. The four-month sample axle spacing data appear to be in great control condition.





**Figure 24: Drive Tandem Axle Spacing Analysis in Average and Standard Deviation Chart of SPC Method at Station 1**



**Figure 25: Drive Tandem Axle Spacing Analysis in Average and Standard Deviation Chart of SPC Method at Station 3**

### 5.3.3 Data Quality Flag

The summation of loop sensor flags is weighted by the number of axle spacing flag type and the corresponding day's traffic volume. This loop sensor performance metric at Stations 1, 2 and 3 is as shown in Figure 26, and the loop sensor plot at Stations 4, 5 and 6 is shown in Figure 27. These data quality flag results from eight loop sensor logic tests gave a better evaluation than four scale sensor logic tests.

In Figure 26, the loop sensor's performance at Station 1 was detected in a bad performance in January and early February. After five weeks of erroneous axle spacing and speed data records, the loop sensor was recalibrated on Feb. 7<sup>th</sup>. At Station 2, the first two weeks, the eighth, ninth and 14<sup>th</sup> week had some errors. At Station 3, the fourth week in March had a small loop sensor flag. Figure 27 clearly represents some periodic small sensor drifts at Stations 4 and 6. Station 4 had a large error in the end of week 13; so did Station 5 in the middle of week 15 and Station 6 in the beginning of week 6.

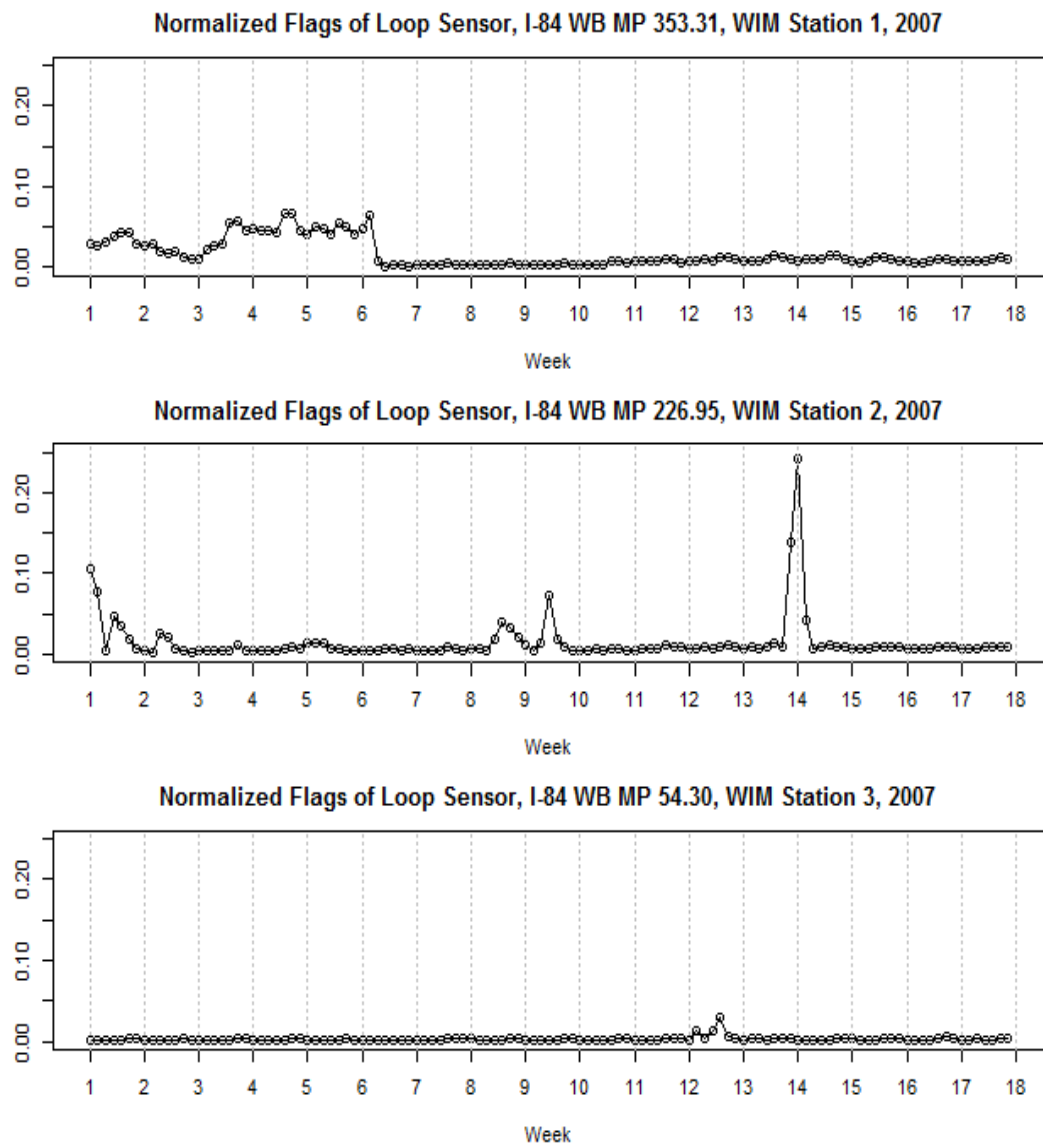


Figure 26: Normalized WIM Weight Sensor Flags at Stations 1, 2 and 3

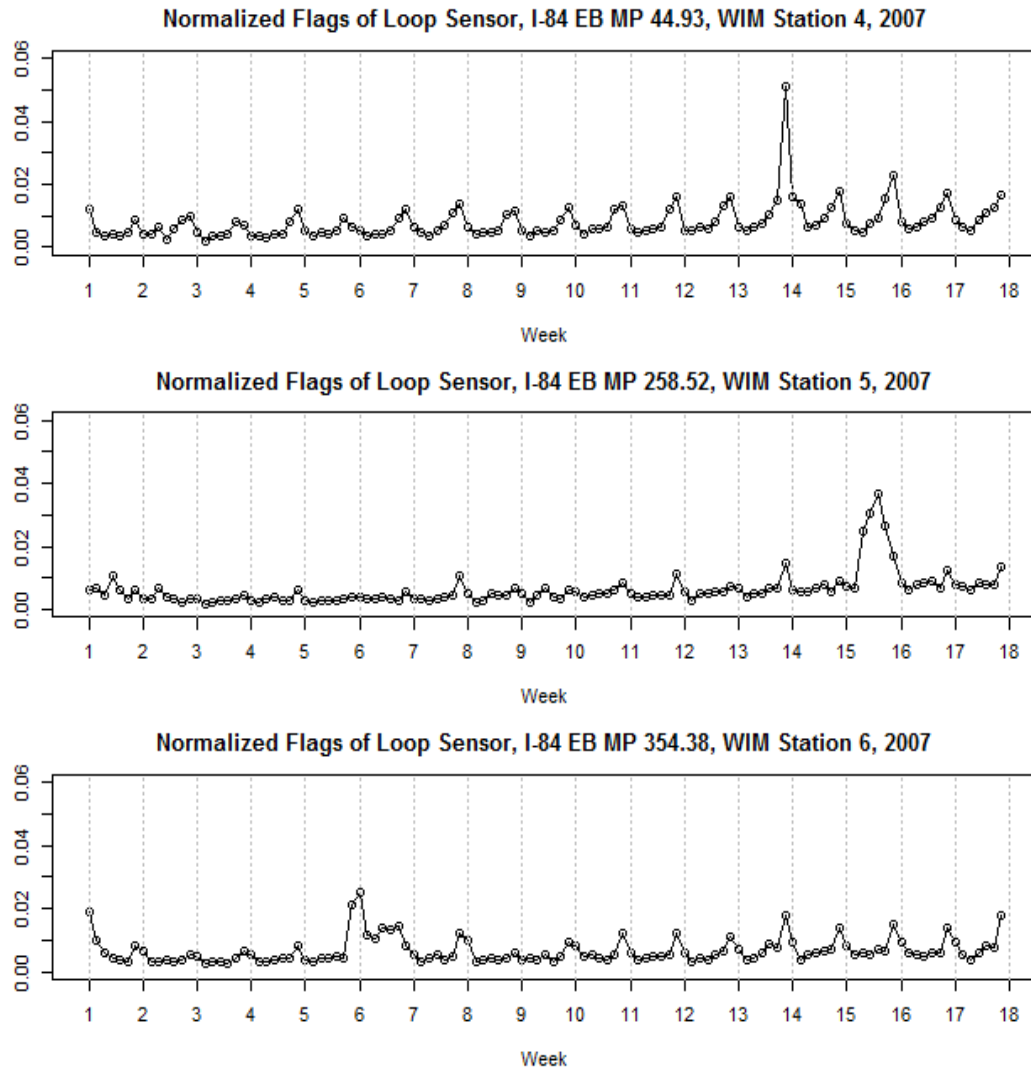


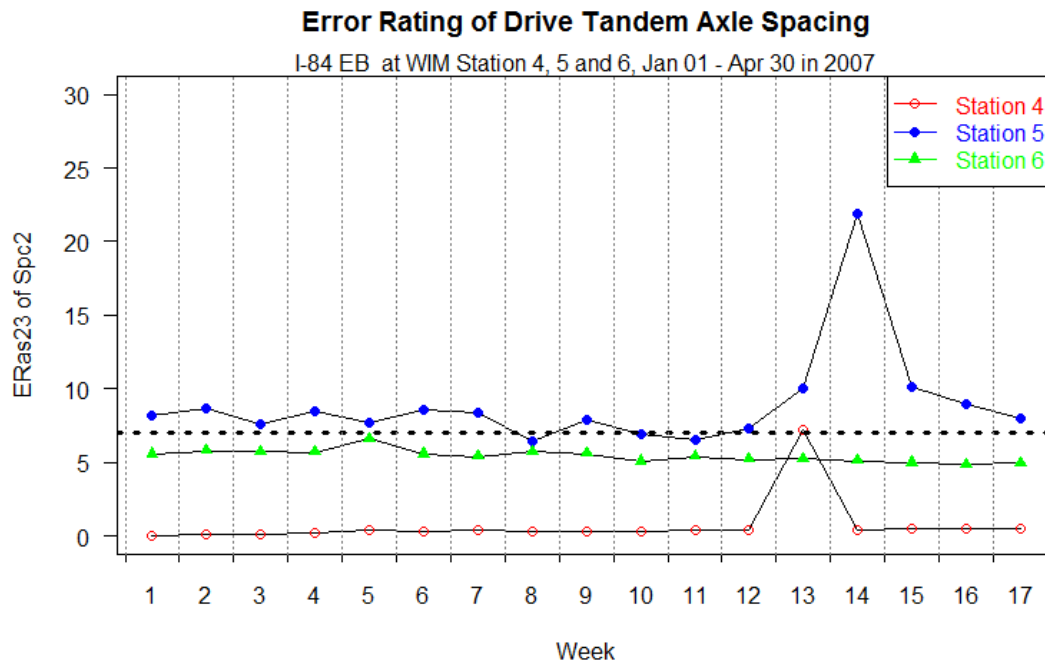
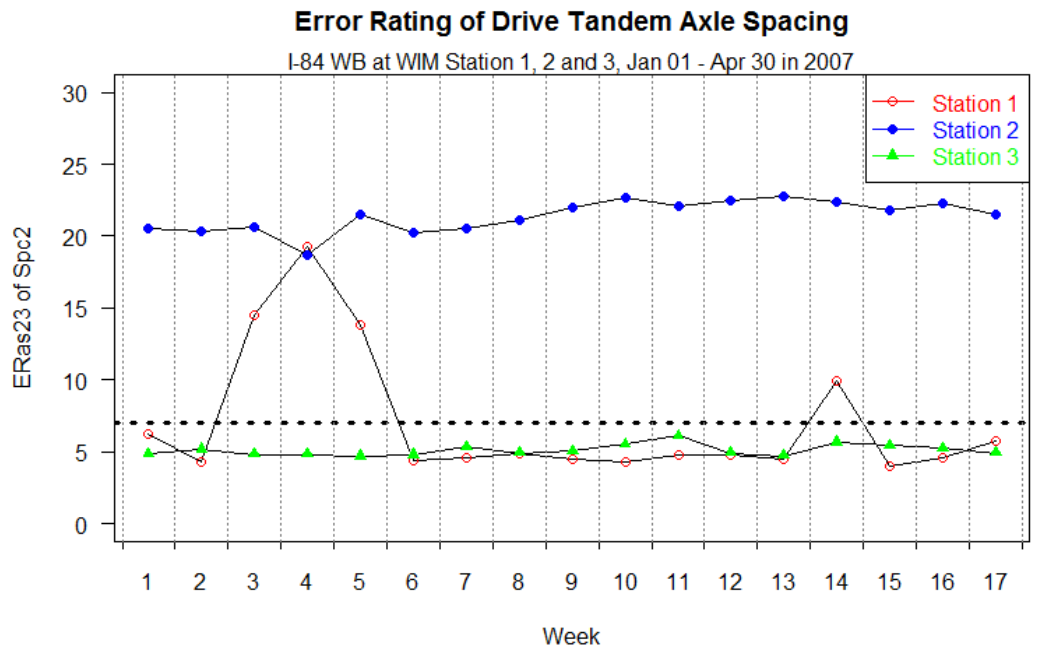
Figure 27: Normalized WIM Weight Sensor Flags at Stations 4, 5 and 6

### 5.3.4 Axle Spacing Error Rating Model

ERas23 values are calculated from a multiple linear regression model developed by Slavik and Wet (2012), which can detect the tendency of a loop

sensor's performance. The ERas23 results using the sample WIM systems are plotted as shown in Figure 28. To facilitate WIM data quality comparison among WIM systems, ERas23 values of WIM Stations 1, 2 and 3 are plotted in the same chart, as are WIM Station 4, 5 and 6 ERas23 values.

Among the six sites, Stations 3 and 4 have the best loop sensor data during this chosen time slice. The ERas23 values of Stations 2 and 5 stay above the maximum acceptable error rating of 7 percent most of the time, which means it's necessary to check and rectify the situation. Station 2 is mostly flat and above the 7 percent rating, suggesting the calibration factor is too high or too low since the algorithm uses absolute average discrepancy. For Station 5, it might be the same reason as with Station 2, or simply because a 7 percent Error Rate is too strict to apply to this station. It is worth noting that this multiple linear regression model is derived on a basis of six-axle and seven-axle trucks instead of five-axle semitrailer trucks. In addition, the interpretation of the graphical results mainly relies on the maximum acceptable error rating of 7. The acceptable rating is concluded from a study that matches the modal results Eras23 and known operating conditions. The objective matching result may be too conservative to apply to Station 5 in this study.



**Figure 28: Error Rating of Drive Tandem Axle Spacing**

## **5.4 Results Comparison**

The following section summarizes these WIM performance results from the data quality methods. To explore the effectiveness of data quality methods, it is necessary to compare their evaluation results in shared measurement criteria. A set of common rating criteria in Section 5.4.1 unifies the evaluation of diverse data quality methods in three levels for reasonable performance-assessed results. In the following sections, the weight data quality and axle spacing data quality results are summarized and discussed. This section is not intended to determine the causes of equipment errors, but to compare the results of data quality methods for axle weight and axle scale measurements.

### **5.4.1 Rating Criteria**

The triangulation method rates the station based on its relative performance within the trip groups of westbound 1-2-3 and eastbound 4-5-6. This is because the principle of the triangulation method is to compare one truck's weight data or axle spacing data at three WIM stations. Therefore, the within-group rating for each station would be more reasonable. A set of weekly triangulation plots is generated to facilitate the quantification of errors as shown



in Appendix D. When the weekly mean variation ratio is a large value shown as a larger dot in the plot, the corresponding rating will shift up a level.

The SPC method rating is based on the Shewhart test results of the SPC  $\bar{X}$  chart. It is also notable that the evaluations of elements on the SPC  $\bar{X}$  chart are weighed by the result of the s-chart. The s-chart checks the shifts in variability and rapid fluctuation. For example, if on the  $\bar{X}$  chart a series of consecutive points is centered on the lower control limit and in good operational condition, the corresponding standard deviations on the s-chart are expected to be around 0 rather than around its central line. For another instance, if the process average on the x-bar chart centered near the center line, it is allowable to have some standard deviation points out of the upper and lower control limit.

The rating of the data quality flag method and finite mixture model distribution method cannot be quantified as an error rating method. Thus, only levels 1 and 3 are used to rate their results. For the error rating method, a level 1 rating is for these below the maximum acceptable error rating of 7 percent, and a level 3 rating is for these above 7 percent.

Triangulation Method:

- Level 1: Least variation ratio counts among the three stations;
- Level 2: Less variation ratio counts among the three stations;
- Level 3: Most variation ratio counts among the three stations.

Statistical Process Control  $\bar{X}$  Chart:

- Level 1: Less than three violations
- Level 2: three or four violations
- Level 3: More than four violations

Data Quality Flag:

- Level 1: Zero normalized flags at the station
- Level 3: Non-zero normalized flags at the station

error rating of Axle Spacing Method:

- Level 1: Less than the maximum acceptable error rating of seven
- Level 3: More than the maximum acceptable error rating of seven

Finite Mixture Model Distribution Method:

- Level 1: No significant departure from the established GVW distribution
- Level 3: Significant departure(s) from the established GVW distribution

#### 5.4.2 Weight Accuracy Results

From the weight data quality comparison results at the six stations, the triangulation method identifies more sensor errors than other methods, and gives signals every time the finite mixture distribution method finds an error. The rating at Station 3 is controversial because the triangulation method has more error scores than other methods. When taking a closer look at the data, steering axle weight at Station 3 did effectively find erroneous weight data: weight data at Station 3 are below 9 kips or above 12 kips, compared to Station 1 and 2's data around 11 kips. The fluctuation of steering axle weight data might be attributed to the temporal changing proportions of the traffic components, as shown in Figure 43. At Station 2, the errors at week 8 were not detected in the triangulation method because the spike of the weekly variation ratio at Station 2 only happened on Wednesday of week 8, while the weekly counts of variation ratios in Stations 1 and 3 are both higher than the counts in Station 2. Other data quality methods evaluate the longitudinal data, in these cases the data at week 8 showed some obvious signals. From weeks 6 to 10, Stations 1 and 3 might have relatively worse weight data quality than Station 2. For Stations 4, 5 and 6, the errors detected by the triangulation method are more than those in other

methods. Sometimes the triangulation method concludes with more conservative results, such as at Station 4. Furthermore, it is worth noting that rating criteria of the triangulation method are relative variations among three stations; by contrast, other data quality methods use the absolute rating estimates that compare data within one station. The relative grading does not provide a quantitative statistics for gaps among different levels. For example, in weeks 3 and 10 the weight data of Station 1 were both rated as the most erroneous data, but in terms of the variation ratio counts difference between levels 2 and 3 in two weeks, there is an immense difference. Hence, a level 3 rating in the triangulation methods could suggest a slight fluctuation in sensor performance or could indicate a severely malfunctioning WIM system. But it is notable that even in good operational condition, the data quality metrics might interpret the system's performance differently due to the vehicle dynamics or pavement profile.

The SPC average control chart and standard deviation chart signals many errors. The SPC concludes with overrated results at Stations 1 and 4, as shown by a wide range of variation ratios in levels 2 and 3. At Station 3, only SPC gives a severe result at weeks 2 and 4. This might be attributed to the central line. The mean of the quality measure – steering axle weight is calculated using the sample

data instead of the accurate data. If the central line moves slightly upwards or downwards, one group of rule 4 violation signals would be absent. This is the same case for Station 4: the true central line might be slightly higher than the current central line so that the rule 4 violations would be detected at Station 4 which will match other data quality methods' results. The SPC method assumes the data are in unacceptable level, ARL is set to be small. This might account why SPC method rates high-level errors more often than other methods.

In the data quality flag plots, only four weight flag rating tests are implemented to diagnose the health of the weight sensor. The data flag plots rate a major SPC is used to monitor and control the whole/part of a process, ensuring it operates at its full potential. The type of SPC chart is dependent upon the type of data collected as well as the subgroup size, as shown by the table below.

Quality of data as good and do not even sense obvious weight scale errors. The other three methods all can detect most flags the data flag method detects (e.g., week 15 of Station 1 and week 6 of Station 6). However the data quality method could not detect the errors found by other methods (e.g., week 2 of Station 4 and week 9 of Station 5).

The mixture distribution methods accurately sense most high variation weight data. The visual GVW component distribution needs subjective interpretations. At Station 6 in the sixth week, the finite mixture method only shows a small fluctuation on the temporary component plot, but on other methods there is a large error in week 6. In data quality flag plots, the level 3 rating is due to checked flags of “GVW  $\pm$  the sum of the axle weights by more than 7%”; in the SPC chart, there is a significant spike during the sixth week; in the triangulation plot, four consecutive points with one large variation point are beyond the control limit. It indicates that GVW may not be an ideal indicator of WIM system accuracy. A GVW-based measurement offsets the positive and negative errors in individual axle load measurement.

**Table 6: WIM Load Cell Scale Data Quality Result Table**

Weight Scale - Station 1					Weight Scale - Station 2					Weight Scale - Station 3				
Week	Tri	SPC	Flag	Mix	Week	Tri	SPC	Flag	Mix	Week	Tri	SPC	Flag	Mix
1	2	2	1	3	1	2	2	1	1	1	1	1	1	1
2	2	1	1	1	2	3	3	1	3	2	1	3	1	1
3	3	3	1	3	3	2	1	1	1	3	1	1	1	1
4	3	2	1	3	4	2	1	1	1	4	1	3	1	1
5	2	3	1	3	5	3	2	1	1	5	1	1	1	1
6	3	2	1	3	6	1	1	1	1	6	2	1	1	1
7	3	2	1	3	7	1	1	1	1	7	2	2	1	1
8	3	2	1	1	8	1	3	3	3	8	2	1	1	1
9	3	2	1	1	9	1	2	1	3	9	2	1	1	1
10	3	3	1	3	10	1	1	1	1	10	2	1	1	1
11	1	2	1	3	11	2	1	1	1	11	2	1	1	1
12	2	2	1	1	12	2	1	1	1	12	3	2	1	1
13	1	2	1	1	13	3	1	1	1	13	2	1	1	1
14	2	2	1	1	14	3	1	1	3	14	1	1	1	1
15	3	3	3	3	15	1	3	1	1	15	2	1	1	1
16	1	2	1	1	16	2	3	1	1	16	3	1	1	1
17	1	2	1	1	17	2	3	1	1	17	3	1	1	1

Weight Scale - Station 4					Weight Scale - Station 5					Weight Scale - Station 6				
Week	Tri	SPC	Flag	Mix	Week	Tri	SPC	Flag	Mix	Week	Tri	SPC	Flag	Mix
1	1	1	1	1	1	3	3	1	3	1	2	1	1	1
2	3	3	1	3	2	3	3	1	3	2	1	3	1	1
3	2	3	1	1	3	3	1	1	1	3	2	3	1	1
4	2	1	1	1	4	3	1	1	1	4	2	3	1	1
5	1	1	1	1	5	3	1	1	1	5	2	1	1	1
6	2	1	1	1	6	1	1	1	1	6	3	3	3	3
7	1	1	1	1	7	2	1	1	1	7	1	1	1	1
8	2	1	1	1	8	3	1	1	3	8	1	2	1	1
9	1	1	1	1	9	3	3	1	3	9	3	1	1	1
10	1	3	1	1	10	3	1	1	1	10	3	1	1	3
11	1	2	1	1	11	2	1	1	3	11	2	1	1	1
12	1	3	1	1	12	2	1	1	3	12	2	1	1	1
13	2	3	1	1	13	3	3	1	3	13	1	1	1	3
14	3	1	1	3	14	2	3	1	3	14	1	1	1	3
15	1	1	1	1	15	3	3	1	3	15	2	3	1	3
16	1	3	1	1	16	2	3	1	1	16	1	1	1	3
17	1	3	1	1	17	3	3	1	1	17	2	1	1	3

### 5.4.3 Axle Spacing Results

In the evaluation of axle spacing measurement, at Station 2 the SPC method concludes with completely different results with the error rating method and the triangulation method. The SPC algorithms could account for this result.

The SPC method requires accurate data for quality evaluation, which could be obtained during the first two weeks after loop sensor calibration in this case. Due to the failure to obtain the calibration schedule to acquire accurate data, the central line and three-sigma line are calculated from the sample data themselves. By contrast, the triangulation method and error rating method both perform the evaluation with the consideration of true distribution of drive tandem axle weight. Thus these two methods can detect the actual facts that the loop sensor at Station 2 is completely drifted off from the process mean. Since the SPC only compares the data with itself, it cannot detect this sensor drift before the evaluation starts and thereby provided a wrong answer after comparing with itself. In addition, the error rating method is to evaluate the data in the context of true variation of drive tandem axle spacing in a practical application so that it also easily finds of loop sensor drift at Station 2. The triangulation method's score is a horizontal between-stations analysis of estimates derived from the 1-2-3 WIM station group and 4-5-6 WIM station group. It is particularly good at detecting these kinds of severe errors.

The SPC method detects more errors at Station 6 than the triangulation method and error rating method, as shown in Table 7. The SPC plot of Station 6



in Figure 41 shows there are two groups of large departure from the drive tandem axle spacing spectra during the 17 weeks. These large deviances significantly increase the process mean, which causes most points to be read as Shewhart rule 4 violations. At Station 5, the SPC method apparently sensed fewer errors than the triangulation method. Even though the selected two classic Shewhart tests effectively detect discernible deviances in the process mean, they are not sensitive to smaller changes such as a one-sigma or two-sigma shift in the process mean. When accurate data are not available as a reference, the SPC method is especially not recommended for WIM data quality analysis.

The data flag method in the drive tandem axle spacing metric is much better than its application in the steering axle weight metric. Among the loop sensor six station's tables, the data flag method almost finds the most severe errors that other methods find and sometimes even senses more errors (e.g., Station 4). At Station 3, the data flag method is the only method that signaled out the errors at week 12. The error marks on the 12<sup>th</sup> week are due to three consecutive days' violations in the check of "any spacing is less than 3.4 feet." The rule-based quality method examines all Class 9 truck axles' weight and

spacing, and thereby flagged the errors that cannot be detected by other axle spacing data methods.

The error rating method is able to detect errors effectively as shown in the comparison Table 7. But in Stations 4 and 6, the Error Rate method is the only method that detects a few data quality errors. The error rating method evaluates all axle sensor performance in the context of entire six- and seven-axle truck populations. The maximum acceptable error rating of 7 might be too conservative when applying it to Oregon's WIM system. An appropriate maximum acceptable error rating data would be good to develop, but would need results from other data quality methods to support it. Overall, when the true average and spread of drive tandem axle spacing - which is decisive parameters of the model - is available from the truck manufacturers' specifications, the axle spacing method is a reliable data quality model to employ.

**Table 7: WIM Loop Sensor Data Quality Result Table**

Loop Sensor - Station 1					Loop Sensor - Station 2					Loop Sensor - Station 3				
Week	Tri	SPC	Flag	Err	Week	Tri	SPC	Flag	Err	Week	Tri	SPC	Flag	Err
1	1	1	3	1	1	3	2	3	3	1	1	1	1	1
2	1	2	3	1	2	3	1	3	3	2	1	1	1	1
3	1	1	3	3	3	3	1	1	3	3	1	1	1	1
4	3	1	3	3	4	3	1	1	3	4	1	1	1	1
5	3	3	3	3	5	2	1	1	3	5	1	1	1	1
6	3	2	3	1	6	3	1	1	3	6	1	1	1	1
7	1	1	1	1	7	3	3	1	3	7	1	1	1	1
8	1	1	1	1	8	3	1	3	3	8	1	1	1	1
9	1	1	1	1	9	3	1	3	3	9	1	1	1	1
10	1	1	1	1	10	3	1	1	3	10	1	1	1	1
11	1	1	1	1	11	3	1	1	3	11	1	1	1	1
12	1	1	1	1	12	3	1	1	3	12	1	1	3	1
13	1	1	1	1	13	3	1	1	3	13	1	1	1	1
14	1	2	1	3	14	3	1	3	3	14	1	1	1	1
15	1	3	1	1	15	3	1	1	3	15	1	1	1	1
16	1	1	1	1	16	3	1	1	3	16	1	1	1	1
17	1	3	1	1	17	3	1	1	3	17	1	1	1	1

Loop Sensor - Station 4					Loop Sensor - Station 5					Loop Sensor - Station 6				
Week	Tri	SPC	Flag	Err	Week	Tri	SPC	Flag	Err	Week	Tri	SPC	Flag	Err
1	1	1	3	1	1	3	2	3	3	1	2	3	3	1
2	2	2	3	1	2	3	1	1	3	2	1	3	1	1
3	3	1	3	1	3	2	1	1	3	3	1	3	1	1
4	2	3	3	1	4	3	1	1	3	4	1	3	1	1
5	1	1	3	1	5	2	1	1	3	5	3	3	3	3
6	1	1	3	1	6	3	2	1	3	6	1	3	3	1
7	2	1	3	1	7	3	2	3	3	7	1	3	3	1
8	2	1	3	1	8	1	1	1	1	8	1	2	3	1
9	2	1	3	1	9	2	1	1	3	9	1	2	1	1
10	2	1	3	1	10	3	2	1	1	10	1	3	3	1
11	1	1	3	1	11	1	2	3	1	11	2	3	3	1
12	1	1	3	1	12	3	2	1	3	12	2	1	3	1
13	3	1	3	1	13	1	2	3	3	13	2	1	3	1
14	3	1	3	3	14	2	1	1	3	14	1	1	3	1
15	2	1	3	1	15	3	2	3	3	15	1	3	3	1
16	2	1	3	1	16	3	2	1	3	16	1	3	3	1
17	2	1	3	1	17	3	2	1	3	17	1	3	3	1

## **6.0 CONCLUSIONS**

This final chapter is a summary of the main findings. The findings are provided in the context of diverse purposes of WIM data quality evaluation in the first section. The second section provides recommendations for possible future work.

### **6.1 Summary of Findings**

This study discusses and explores the state-of-the-art WIM data quality methods and identifies new methods for WIM data quality evaluation. These methods identify the possible equipment malfunctions, calibration drift, etc. Improving the accuracy of truck weighing systems may increase the number of overweight trucks detected. The diverse data quality methods are supported by the robust data quality metrics - steering axle weight and drive tandem axle spacing.

The triangulation method tracks the trucks traveling among the WIM systems at a corridor level, and evaluates WIM systems relative performance in the quality measures - steering axle weight and drive tandem axle spacing. Most existing methods only compare one WIM system's performance at different timestamps. This triangulation method does not employ the longitudinal data as

other research does, but analyzes a group of WIM systems' performance in the same day and effectively controls the meteorological variable. The newly identified variation ratio data quality metric is assessed in a reasonable variation range. This method finally identified most possible WIM malfunctions that other methods sensed.

The SPC approach statistically monitors process behavior of weight and axle spacing accuracy, and differentiates between sensor noises and events that require interventions. An apparent shortcoming of SPC is that it does not evaluate the data quality in the context of WIM system operation so that SPC requires the accurate data to support a sound SPC data quality analysis. As proved by the results at Station 3, when the sample data are selected correctly in a slow sensor drift period, the results from this analysis could be completely wrong because it could not detect the slow sensor drifting. Once the reference accurate data are available, the SPC approach should be the first choice to examine WIM data quality. If a high-accuracy WIM data check is required, SPCs exponentially weighted moving average would be better adopted instead of the Shewhart chart. In summary, if the accurate data for quality metrics are accessible, SPC control charts would be the powerful tool for WIM operators to

audit and communicate in practice. The process shifts, out-of-control conditions, or corrective actions can all be detected and noted on the chart.

The data quality flag method is the most basic data quality method among these evaluation methods. It interprets and assesses data quality on a highly broad level. Its results were not accurate, thus its weight and axle spacing logic checks are not necessarily a valid WIM data accuracy method, but the method can be an effective data cleaning approach applied before any kinds of WIM systems application. More weight logic checks have to be developed to improve the method in order to use it as a data quality method.

The finite mixture model is an additional tool to complement the existing data quality methods. It needs subjective interpretations. Testing of the finite mixture model tool employs GVW data instead of steering axle weight data in other methods. This is not an ideal WIM data quality metric because GVW is not consistent for steering axle weight or drive tandem axle spacing. For the same steering axle load GVW may vary by a factor of 2.5 to 3.0, depending on the loading according to the Oregon WIM data. In addition, for the GVW component distribution far from normal, this method works poorly. As proved by the weight sensor comparison table, the finite mixture model's results roughly

matched the triangulation method's results, indicating it can be used as a powerful additional tool to validate the study results.

Error rating of drive tandem axle spacing method could be an efficient and easy WIM data quality approach only if the true drive tandem axle spacing mean and spread is known. It can possibly diagnose a soundness for an entire WIM system depending on the WIM system types (if WIM weight sensor is Class I or II piezoelectric sensor, because both of them measures the time difference parameter for the calculation of drive tandem axle spacing of each truck). The accuracy of an error rating method depends how close the process mean and variation estimates are to true mean and standard deviation of drive tandem axle spacing spread. The results of error rating methods were accurate and corresponded with other methods.

## **6.2 Future Work**

Accurate WIM data are desirable in a multitude of WIM application fields, such as bridge and pavement design, system planning and commercial vehicle weight enforcement. This research applied different methods to evaluate the WIM data quality. It is the base to examine the early detection of system

malfunctions, sensor drift, etc. Hence, there is a need to examine all of the WIM system's performance statewide.

The newly developed triangulation method is a simple and reliable data quality method for examination in the WIM system's performance at a corridor level. If possible, performing a comparison of more than three WIM devices in a trip would be meaningful. This could be a substitution of employing the reasonable variation ratio range to identify the in-control data and determine the malfunctioning WIM station. Additionally, for the data quality flag method, it would be rewarding to develop more weight data quality tests, even if it would be used as a data cleaning tool.

A periodic WIM data quality analysis is the base to achieve good WIM calibration. The external factors are recommended to collect and explore for establishing a solid site database, including truck and station characteristics. Even though the WIM station characteristics are not accessible in this study, it could be incorporated for WIM data quality study in the future. The knowledge of external factors is recommended to be incorporated into the data validation and analysis process. The truck characteristics can be obtained while assigning the transponder, and can be updated and expanded during on-site calibration



and acceptance testing of the WIM equipment. Other truck characteristics include empty vs. loaded trends, seasonal variations, enforcement effects, unique vehicles and traffic operating characteristics. Physical characteristics of a WIM station include pavement condition and profile, grade, traffic flow restrictions and weather. These profiles derived from the site database will be helpful to account for data abnormalities. In addition, the key meteorological data (e.g., moisture, temperature, cross wind and precipitation) can be included in the site database to further enhance data quality control. Depending on the WIM station, this data may be available from the National Oceanic and Atmospheric Administration.

## 7.0 REFERENCES

- Blab, R., & Jacob, B. (2000). Test of a Multiple sensor and Four Portable WIM Systems. *Heavy Vehicle Systems, Vol. 7, No. 2/3*, 111-118.
- Cebon, D., & Winklet, C. (1991). Multiple-Sensor for Counting and Classifying Vehicles. *Transportation Research Record*, 70-78.
- Champ, C., & Woodall, W. (1987). Exact Results for Shewhart Control Charts with Supplementary Runs Rules. *Technometrics*, 393-399.
- Chan, T., & Miao, T. (2005). Statistical Models from Weigh-in-Motion Data. *Structural Engineering and Mechanics*, 85-110.
- Cook, J. (2011, June). Identity Parade. *Traffic Technology International* , pp. 52-62  
<http://viewer.zmags.com/publication/956bceea#/956bceea/1>.
- Dahlin, C. (1992). Proposed Method for Calibrating Weigh-in-Motion Systems and for Monitoring That Calibration Over Time. *Transportation Research Record* 1364, 161-168.
- Elkins, L., & Higgins, C. (2008). *Development of Truck Axle Spectra from Oregon Weigh-in-Motion Data for Use in Pavement Design and Analysis*. Salem, OR: Oregon Department of Transportation.
- Grundy, P., Grundy, J., Khalaf, H., & Casagrande, R. (2002). Accommodation of Time Dependent Drift of WIM Data. *Third International Conference on Weigh-in-Motion (ICWIM3)*. Orlando, Florida .
- Gyenes, L., & Mitchell, C. (1992). *The Spatial Repeatability of Dynamic Pavement Loads Caused by Heavy Goods Vehicles*. London: Heavy vehicles and roads: technology, safety and policy.
- Han, C., Boyd, W., & Marti, M. (1995). Quality Control of Weigh-in-Motion systems Using Statistical Process. *Transportation Research Record* 1501, 72-80.
- Izadmehr, B., & Lee, C. (1987). On-Site Calibration of WIM Systems. *Transportation Research Record* 1123 , 136-144.
- Juran, J. (1974). *Juran's Quality Control Handbook, 4th ed.* New York: McGraw-Hill Book Company.
- Karki, J. (2000). *Signal Conditioning Piezoelectric Sensors*. Texas Instruments.
- Kweon, Y.-J., & Cottrel, B. (2011). Analysis of Weigh-in-Motion Data for Truck Weight Grouping in Mechanistic-Empirical Pavement Design Guide. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2256, 169-78.
- Larsen, D., & McDonnell, A. (1999). *Second Interim Report on the Installation and Evaluation of Weigh-in-Motion Utilizing Quartz Piezo Sensor Technology*. Connecticut Department of Transportation.
- Lu, Q., & Harvey, J. (2006). Characterization of Truck Traffic in California for Mechanistic-Empirical Design. *Transportation Research Record*, No. 1945, 61-72.

- Lu, Q., Zhang, Y., & Harvey, J. (2009). Estimation of Truck Traffic Inputs for M-E Pavement Design in California. *Transportation Research Record No. 2095*, 62-72.
- Materials, A. S. (2002). *Standard Specification for Highway Weigh-in-Motion Systems with User Requirement and Test Method*. ASTM Committee E-17 on Vehicle-Pavement Systems.
- McCall, B., & Vodrazka, W. (1997). *State's Successful Practices Weigh-in-Motion Handbook*. Washington, D.C.: Federal Highway Administration.
- Monsere, C., & Cetin, M. (2011). *Exploratory Methods for Truck Re-identification in A Statewide Network Based on Axle Weight and Axle Spacing Data to Enhance Freight Metrics*. Portland, OR: Oregon Transportation Research and Education Consortium.
- Nichols, A., & Bullock, D. (2004). *Quality Control Procedures for Weigh-in-Motion Data*. West Lafayette, Indiana: Purdue University Purdue e-Pubs.
- Nichols, A., & Bullock, D. (2005). Evaluation of Drive Tandem Axle Weights for Assessing Weigh-in-Motion Accuracy. *Transportation Research Record*.
- Nichols, A., & Cetin, M. (2007). Numerical Characterization of Gross Vehicle Weight Distributions from Weigh-in-Motion Data. *Transportation Research Record. Journal of the Transportation Research Board, No. 1993*, 148-154.
- Nichols, A., Bullock, D., & Schneider, W. (2009). Detecting Differential Drift in Weigh-in-Motion Wheel Track Sensors. *Transportation Research Record 2121*, 135-144.
- Oakland, J. (1996). *Statistical Process Control, 3rd Edition*. Oxford, England: Butterworth-Heinemann.
- Oman, M. (2010). MnROAD Traffic Characterization for the Mechanistic-Empirical Pavement Design Guide using Weigh-in-Motion Data. *Transportation Research Record*.
- Ott, W., & Papagiannakis, A. (1996). Weigh-in-Motion Data Quality Assurance Based on 3-S2 Steering Axle Load Analysis. *Transportation Research Record 1536*, 12-18.
- Papagiannakis, A., Bracher, M., & Jackson, N. (2006). Utilizing Clustering Techniques in Estimating Traffic Data Input for Pavement Design. *Journal of Transportation Engineering No. 132*, 872-879.
- Papagiannakis, A., Senn, K., & Huang, H. (1996). On-Site Calibration Evaluation Procedures for WIM Systems. *Transportation Research Record 1536*, 1-11.
- Pelphrey, J., Higgins, C., Sivakumar, B., & Groff, R. (2008). State-Specific LRFR Live Load Factors Using Weigh-in-Motion Data. *Journal of Bridge Engineering, Vol.13, No.4*, 339-350.
- Pierchala, C., & Surti, J. (1999). *Control Charts as a Tool in Data Quality Improvement*. Washington, D.C.: National Highway Traffic Safety Administration (NHTSA) Technical Report No. DOT HS 809 005.
- Prozzi, J., & Hong, F. (2007). Effect of Weigh-in-Motion System Measurement Errors on Load-Pavement Impact Estimation. *Journal of Transportation Engineering*.

- Quinley, R. (2009). *Weigh-in-Motion Data Analyst's Manual*. Federal Highway Administration Office of Pavement Technology.
- Ramachandran, A., Taylor, K., Stone, J., & Sajjadi, S. (2011). NCDOT Quality Control Methods for Weigh-in-Motion Data. *Public Works Management and Policy*.
- Sayyady, F., Stone, J., Taylor, K., Jadoun, F., & Kim, R. (2009). Using Clustering Analysis to Characterize MEPDG Traffic Data in North Carolina. *Transportation Research Record*.
- Scholz, M. (2012, 1). *Pavement Structures*. Retrieved 2 15, 2013, from [http://classes.engr.oregonstate.edu/cce/winter2012/ce492/Modules/04\\_design\\_parameters/wim.htm](http://classes.engr.oregonstate.edu/cce/winter2012/ce492/Modules/04_design_parameters/wim.htm)
- Slavik, M. (May 2008). *Weigh-in-Motion Accuracy Verification Through Simulation*. Paris: International Conference on Heavy Vehicles.
- Slavik, M. (September, 2011). Using Weigh-in-Motion Measurements. *South African Institution of Civil Engineering Vol. 19, No. 8*.
- Slavik, M., & Wet, G. D. (2012). Checking WIM Axle-Spacing Measurements. *International Conference Weigh-in-Motion 6*. Fairmont Dallas, TX.
- Smith, B., & Diefenderfer, B. (2010). Analysis of Virginia-Specific Traffic Data for Use with the Mechanistic-Empirical Pavement Design Guide. *Transportation Research Record*.
- Southgate, H. F. (1999). *Quality Assurance of Weigh-in-Motion Data*. Federal Highway Administration.
- Taylor, B., & Bergan, A. (1993). *The Use of Dual Weighing Elements (Double Threshold) to Improve the Accuracy of Weigh-in-Motion Systems, and the Effect of Accuracy on Weigh Station Sorting*. Oregon Department of Transportation, under Contract for the Port of Entry Advanced Sorting System (PASS) Project.
- Turochy, R., & Smith, B. (2002). Measuring Variability in Traffic Conditions Using Archived Traffic Data. *Transportation Research Record 1804*, 168-172.

## 8.0 APPENDIX

### APPENDIX A: Truck Characteristics Analysis

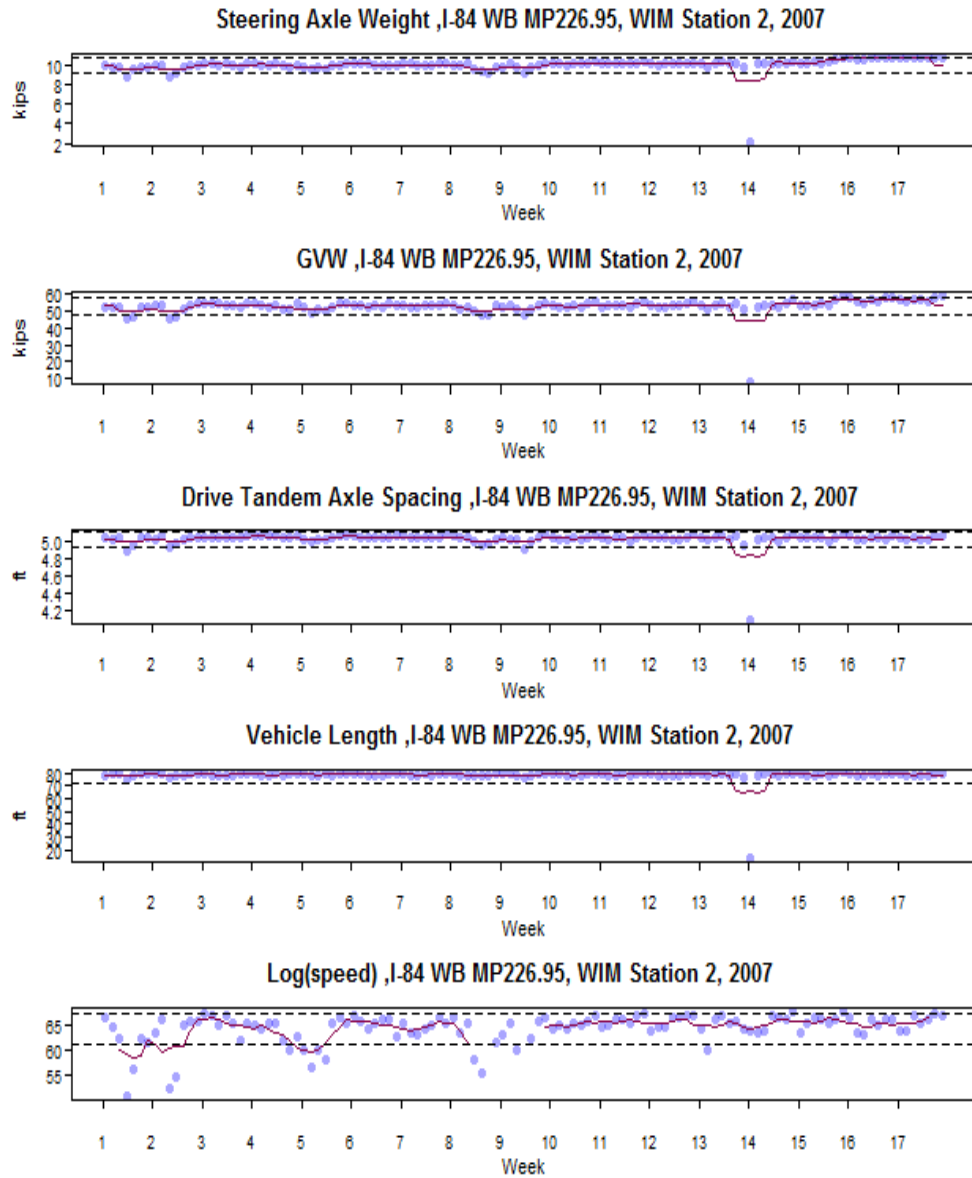


Figure 29: Time Series Moving Average plot at WIM station 2

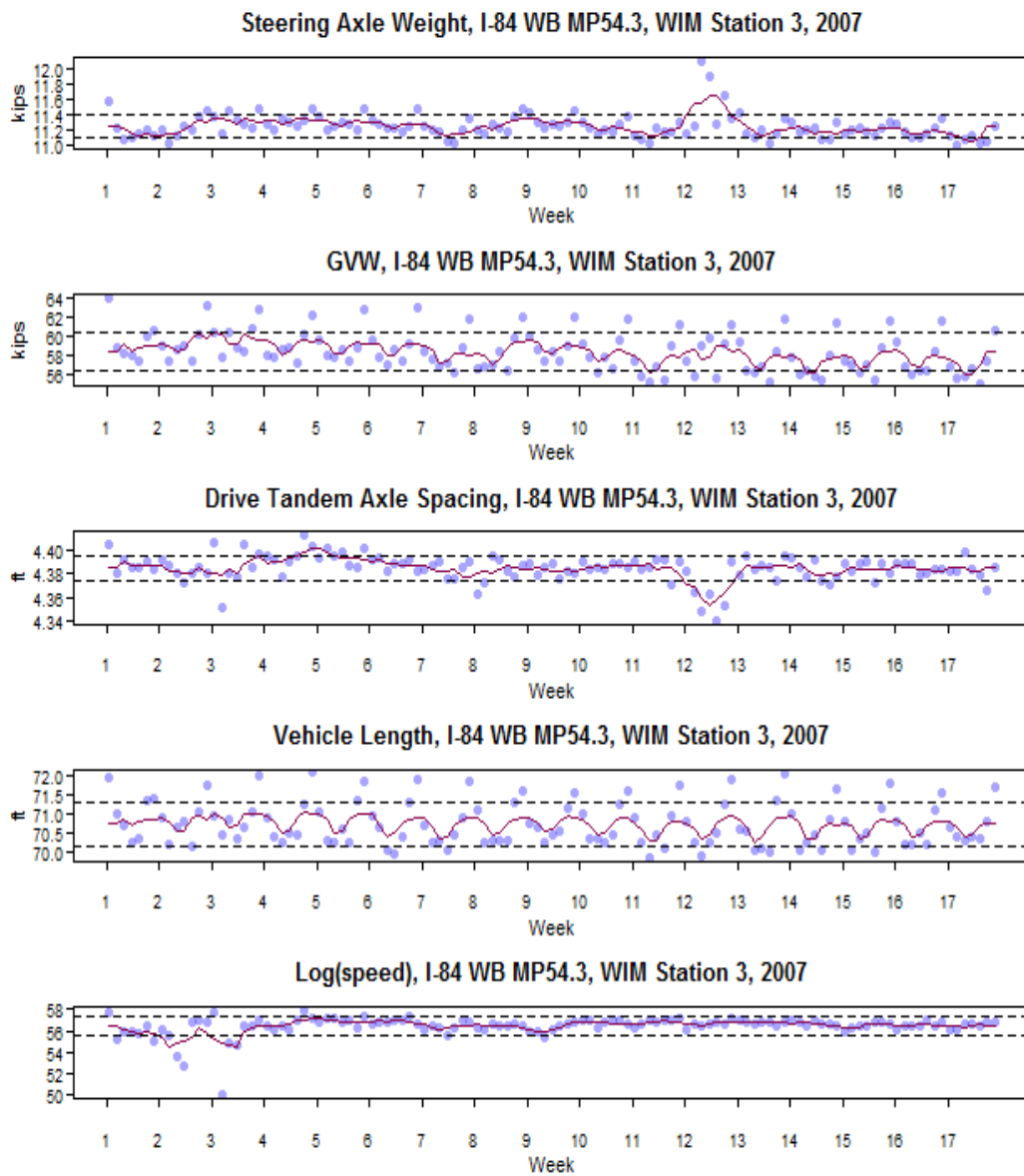


Figure 30: Time Series Moving Average plot at WIM station 3

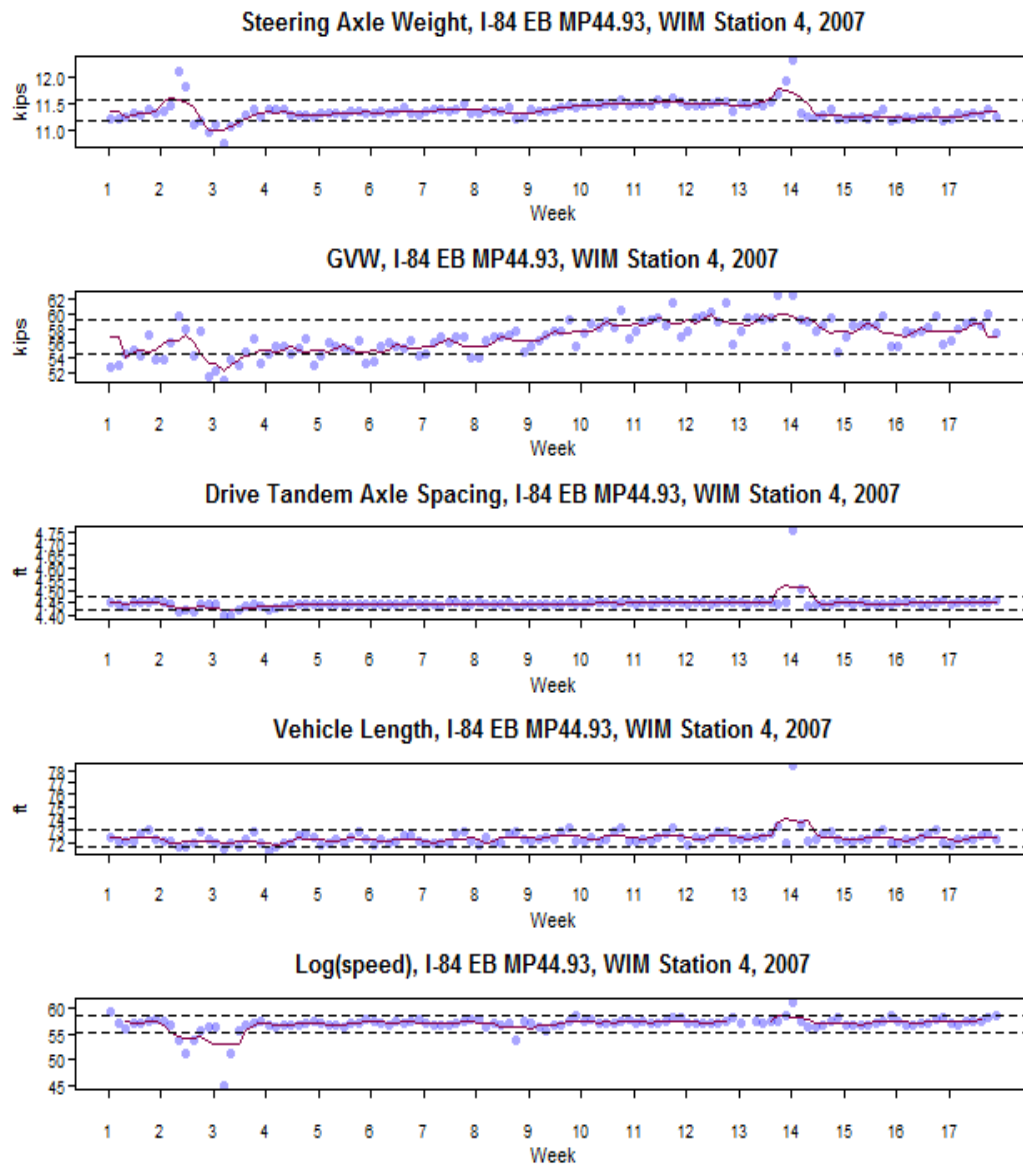


Figure 31: Time Series Moving Average plot at WIM station 4

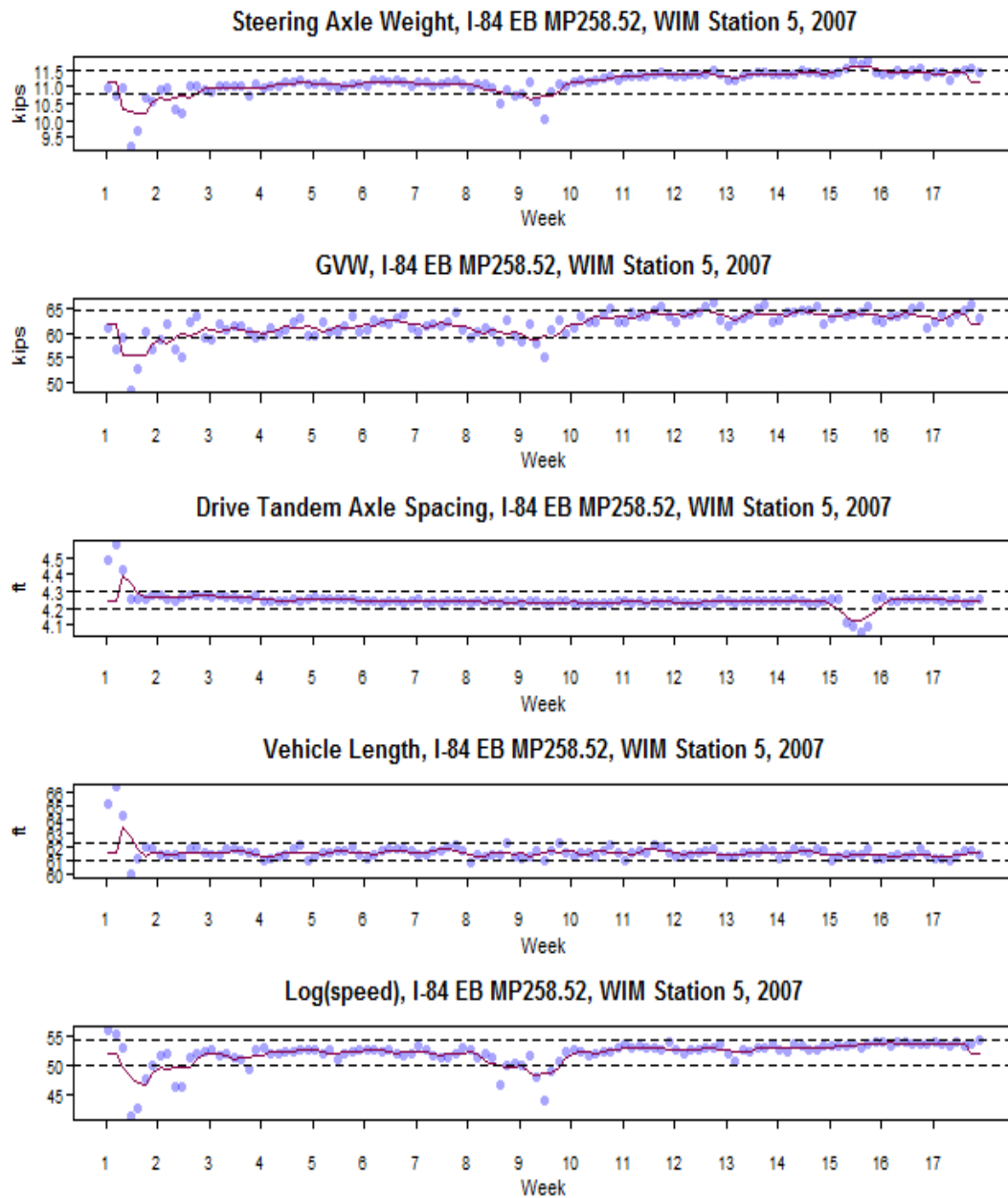


Figure 32: Time Series Moving Average plot at WIM station 5



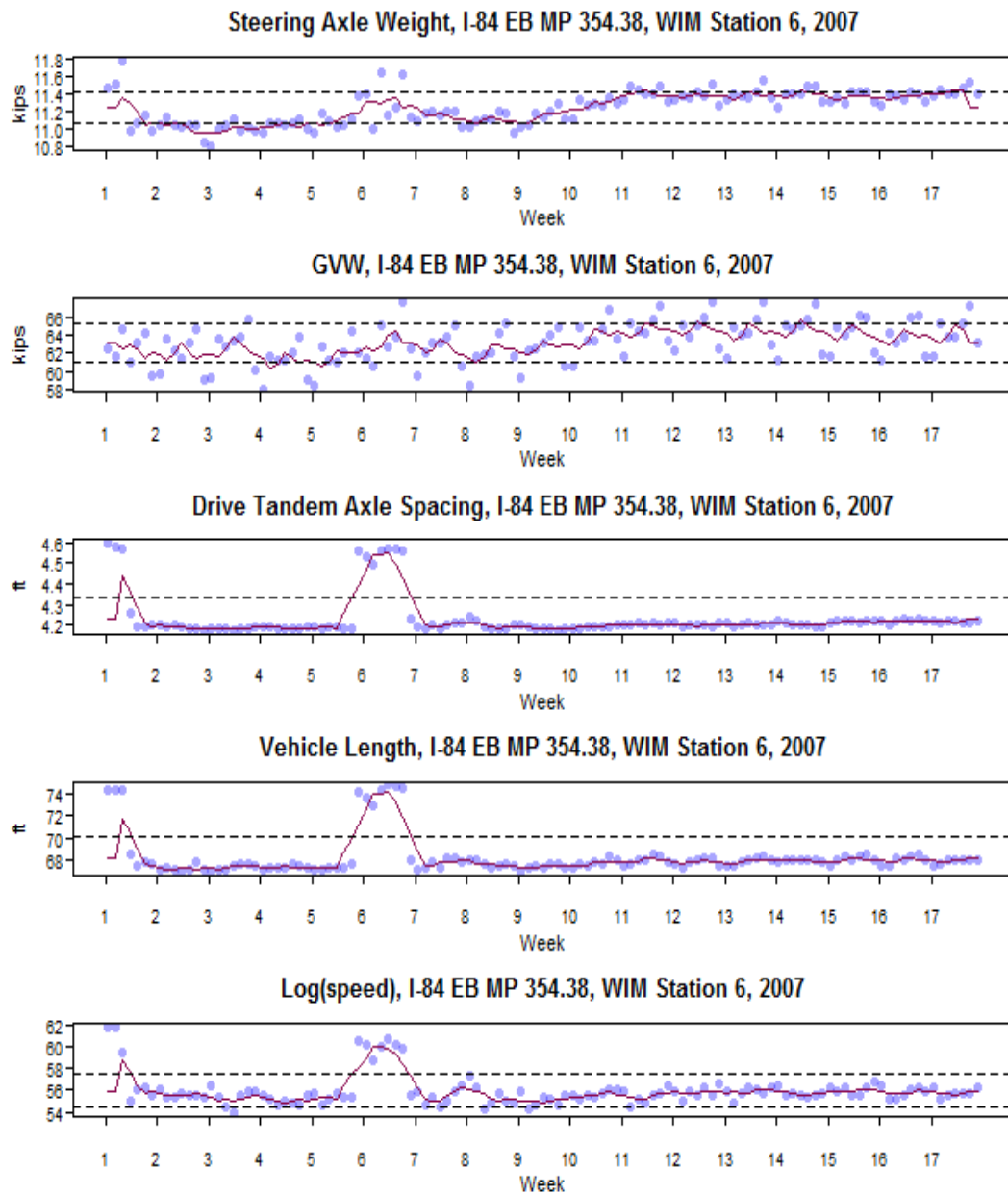
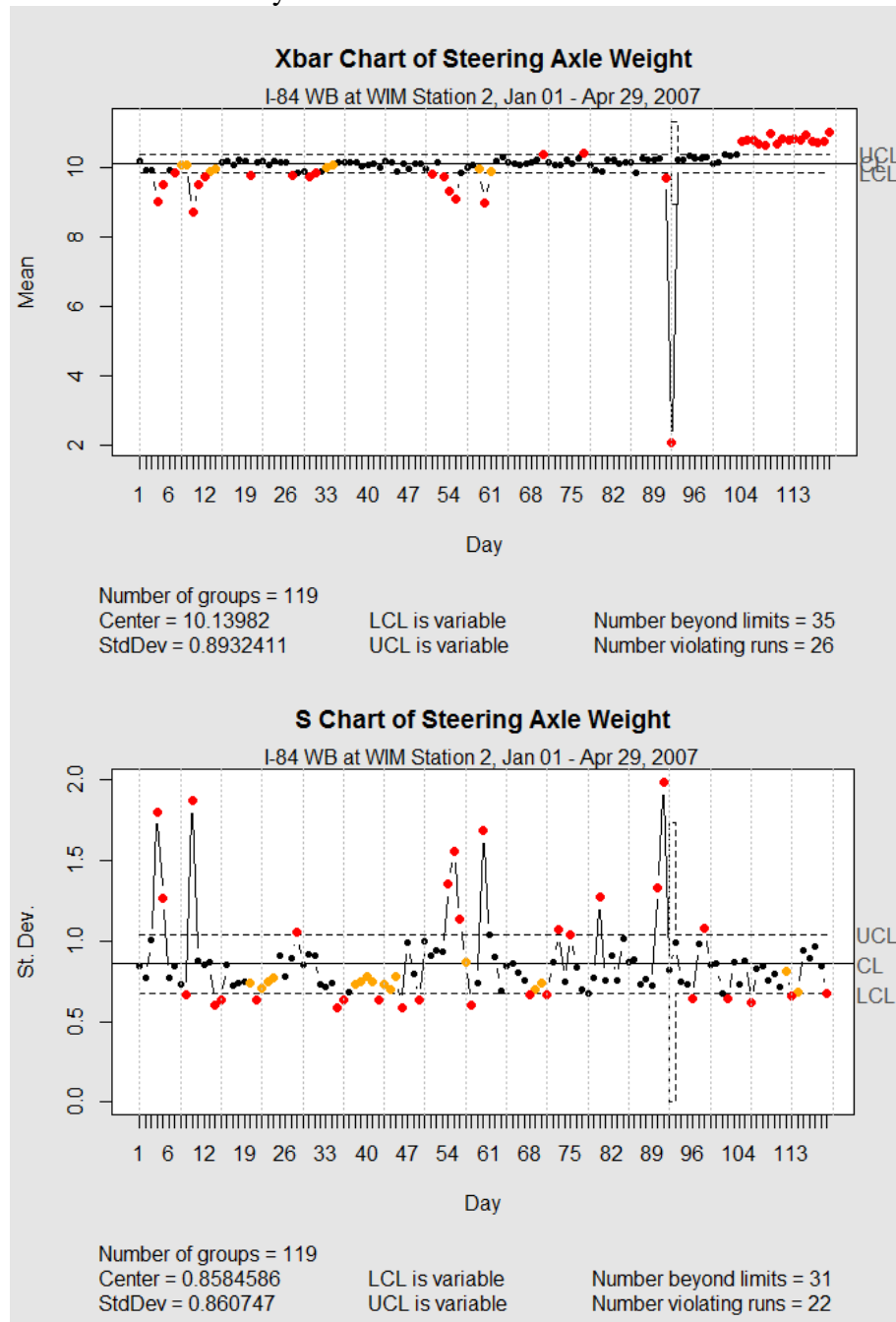
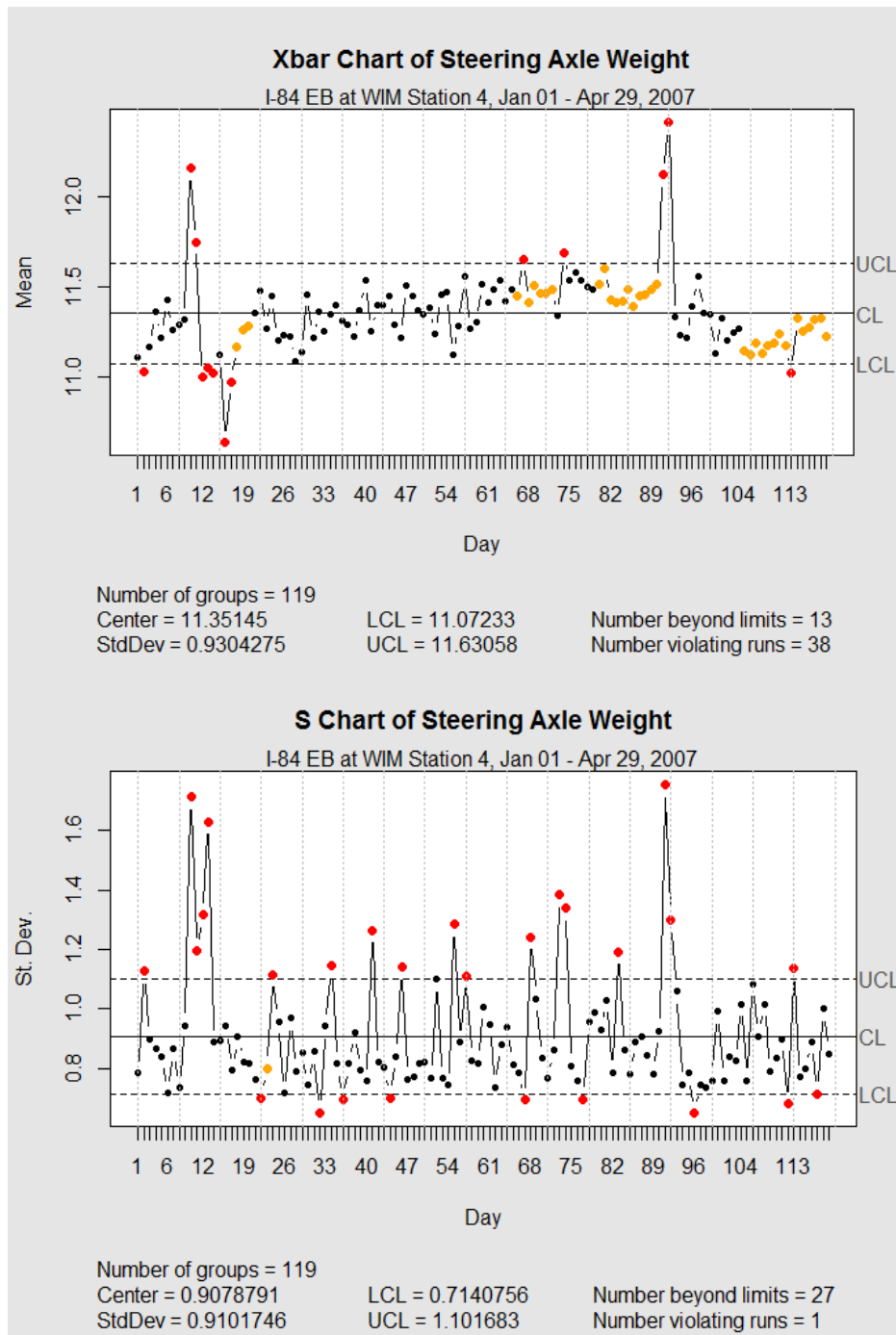


Figure 33: Time Series Moving Average plot at WIM station 6

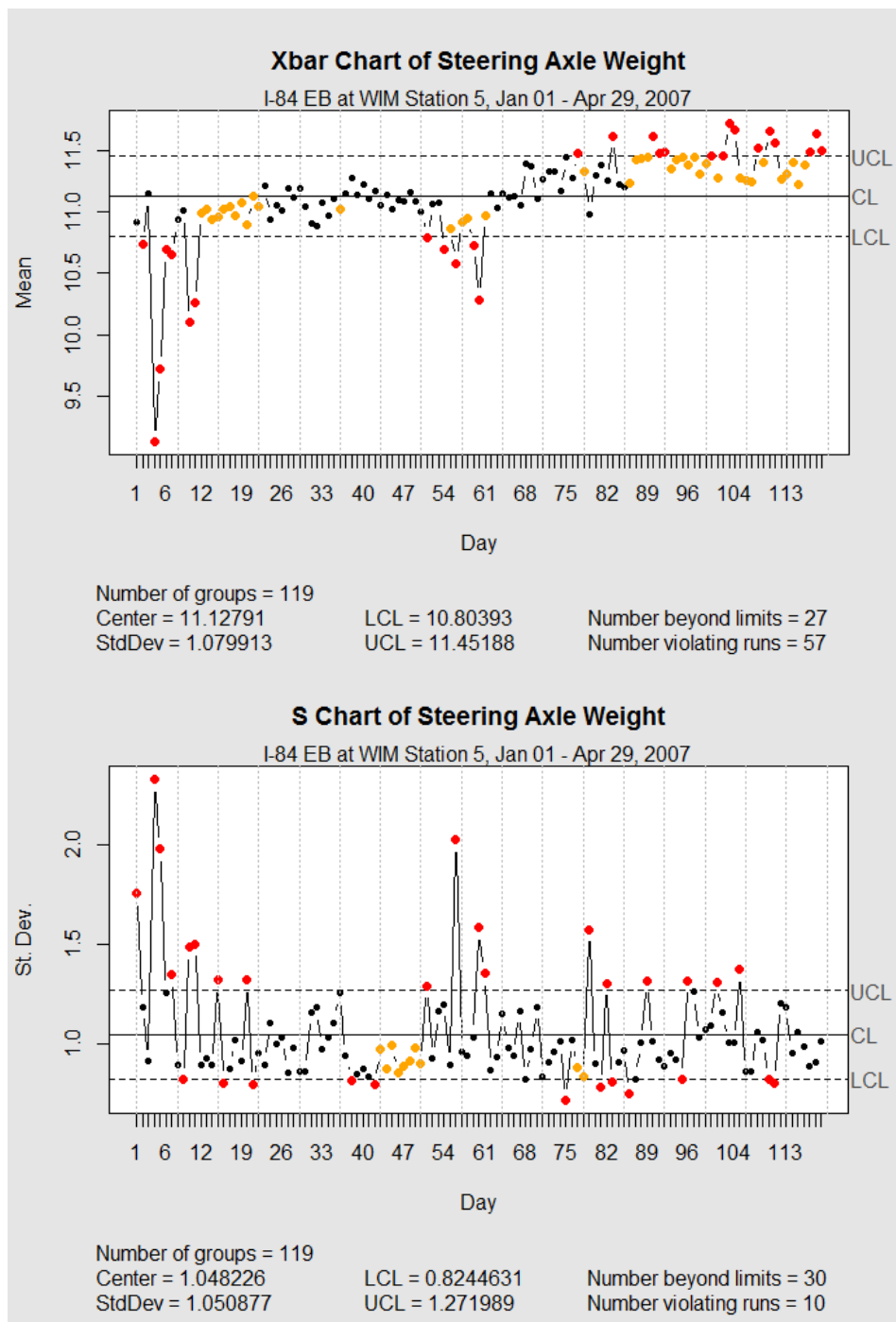
## APPENDIX B: Analysis in Statistical Process Control



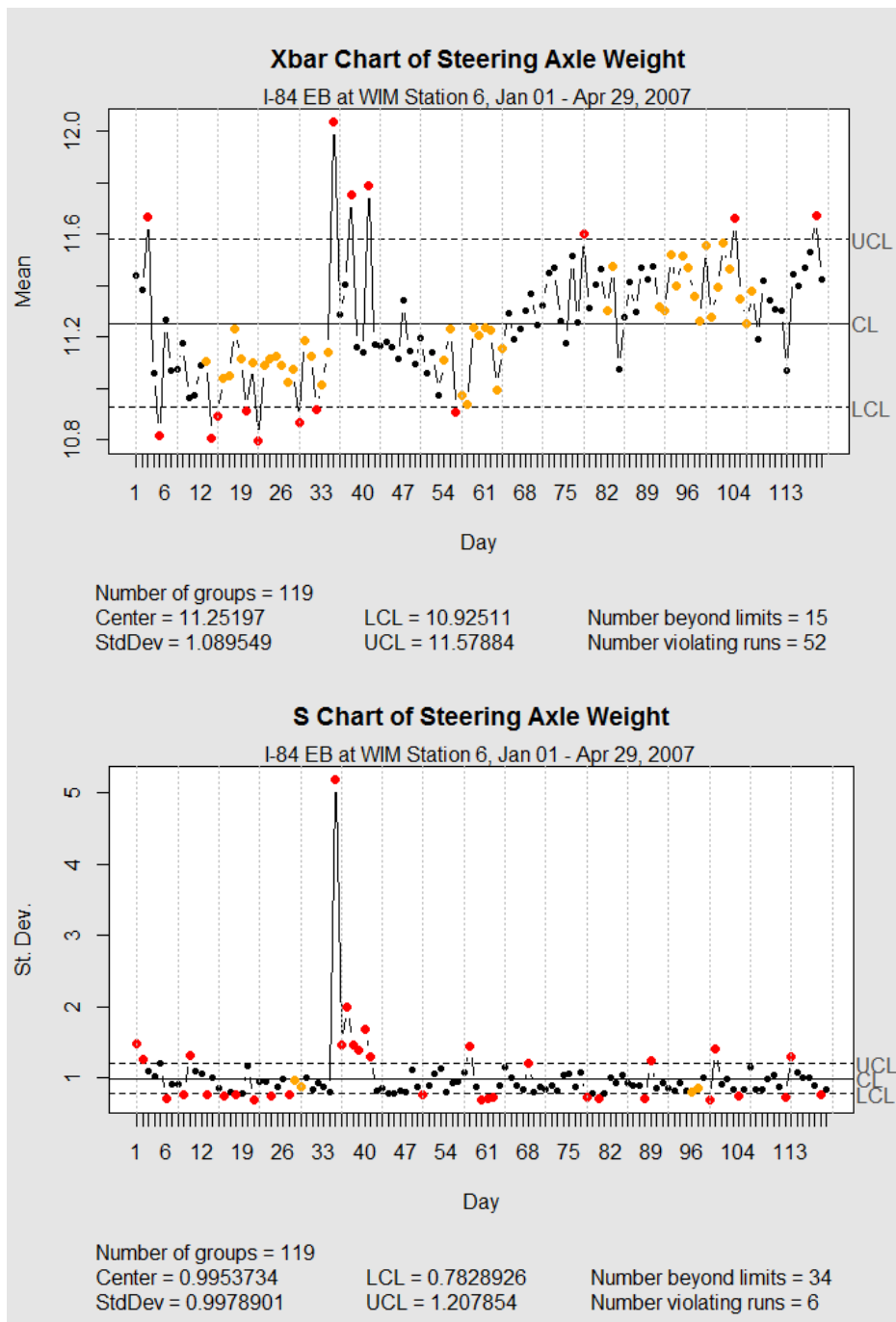
**Figure 34: Steering Axle Weight Analysis in Average and Standard Deviation Chart of SPC Method at Station 2**



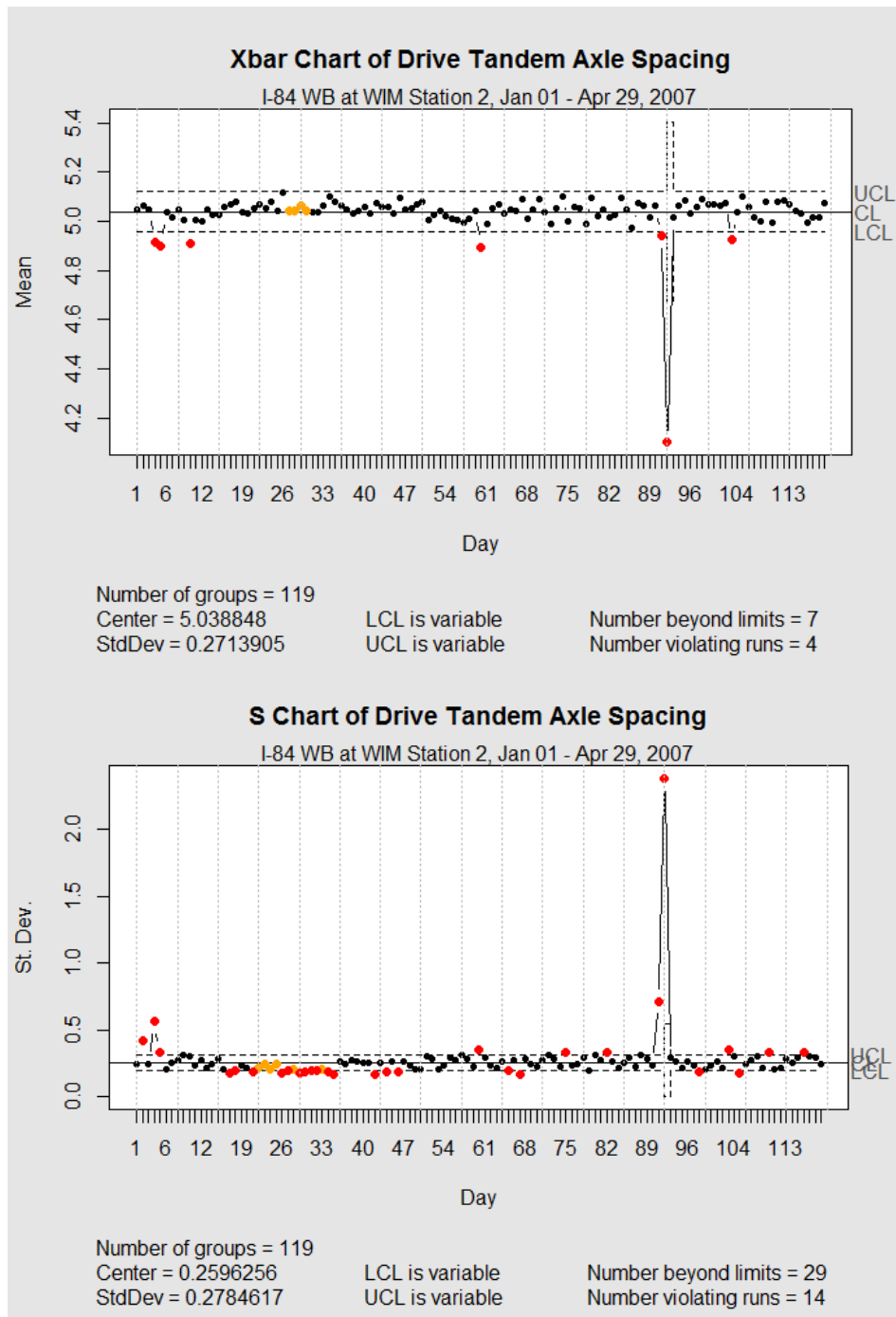
**Figure 35: Steering Axle Weight Analysis in Average and Standard Deviation Chart of SPC Method at Station 4**



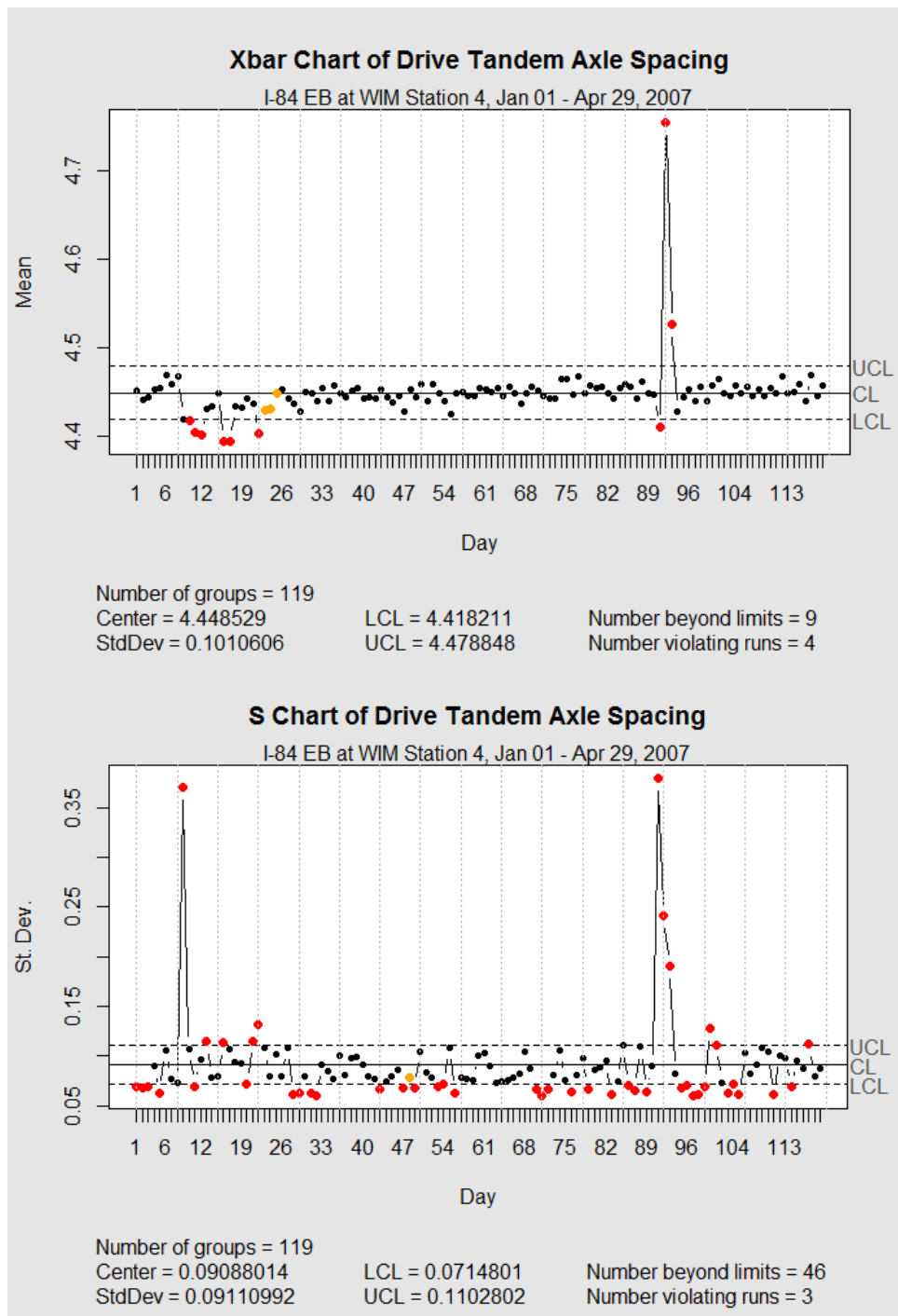
**Figure 36: Steering Axle Weight Analysis in Average and Standard Deviation Chart of SPC Method at Station 5**



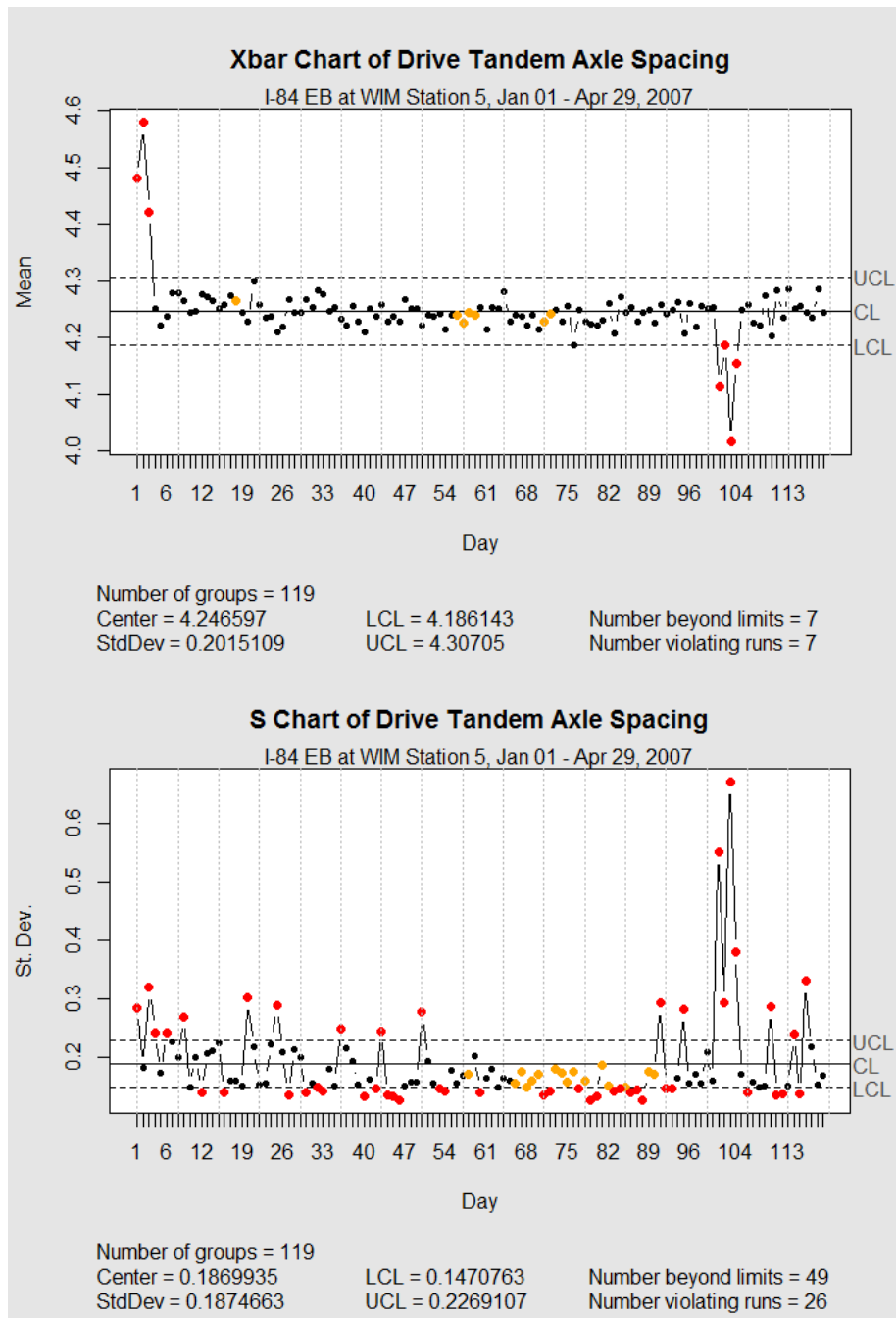
**Figure 37: Steering Axle Weight Analysis in Average and Standard Deviation Chart of SPC Method at Station 6**



**Figure 38: Drive tandem axle spacing Analysis in Average and Standard Deviation Chart of SPC Method at Station 2**

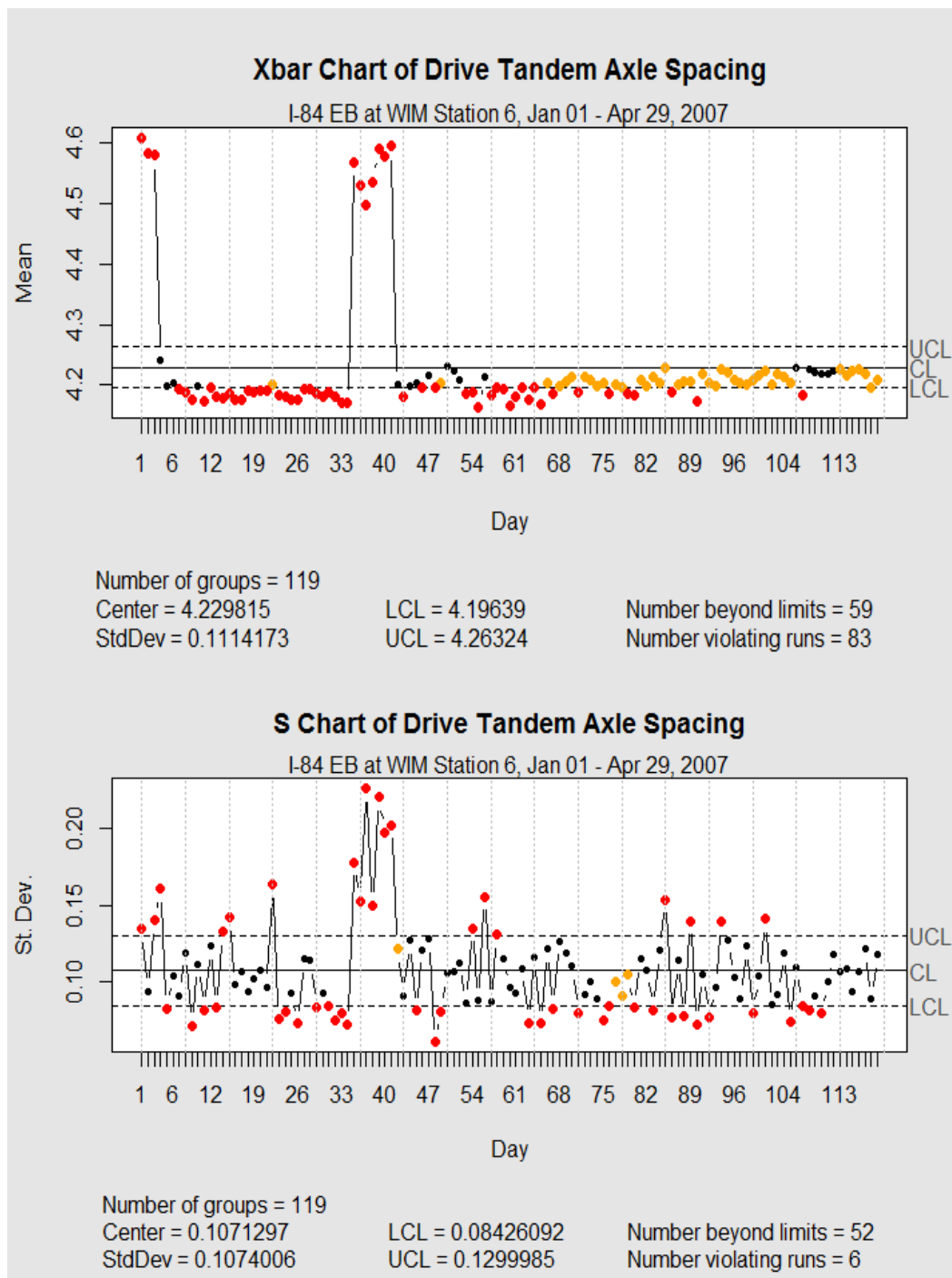


**Figure 39: Drive tandem axle spacing Analysis in Average and Standard Deviation Chart of SPC Method at Station 4**



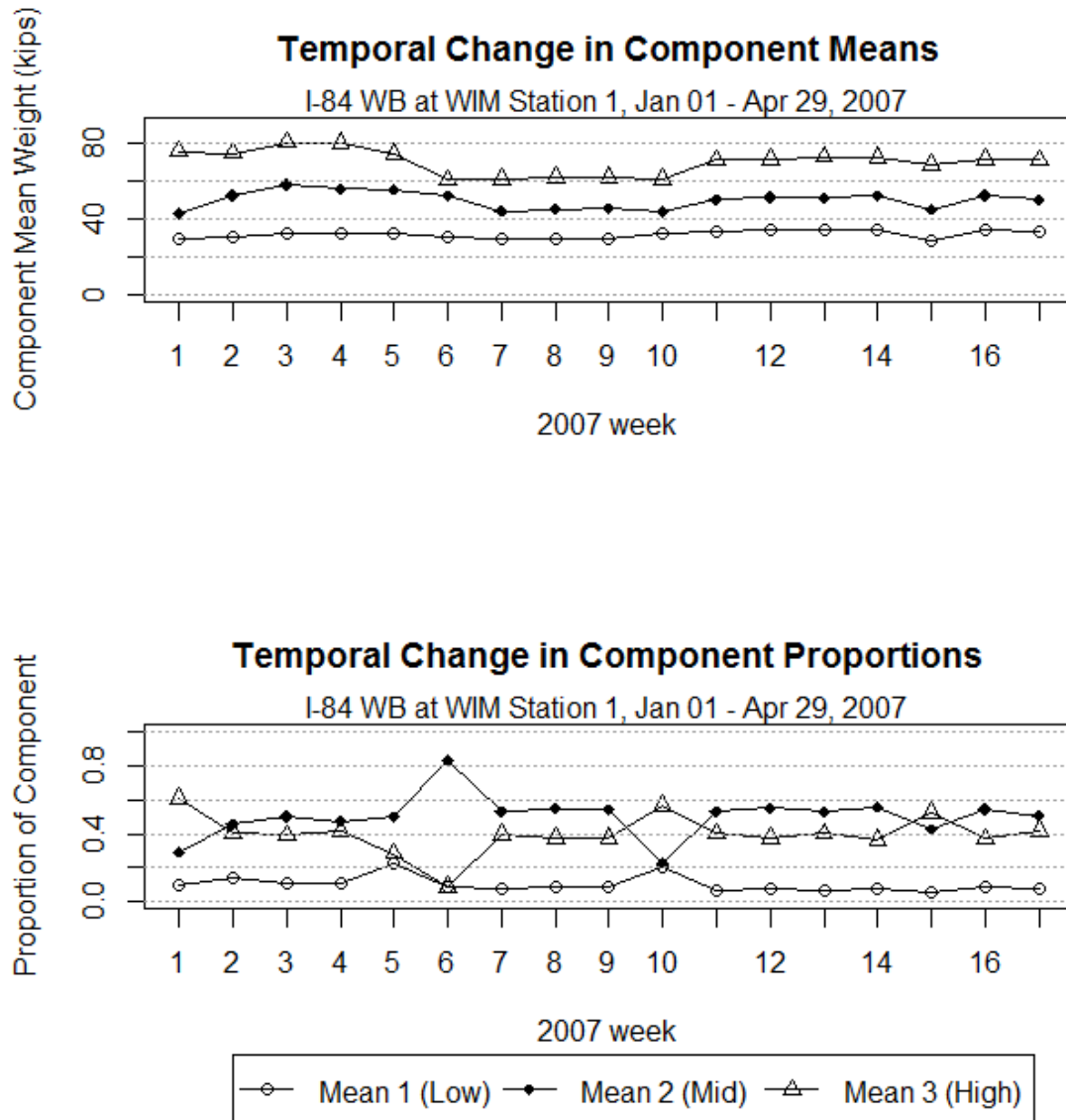
**Figure 40: Drive tandem axle spacing Analysis in Average and Standard Deviation Chart of SPC Method at Station 5**





**Figure 41: Drive tandem axle spacing Analysis in Average and Standard Deviation Chart of SPC Method at Station 6**

## APPENDIX C: Mixture Model Temporal Changes



7

Figure 42: Component Mean and Proportions in WIM station 1

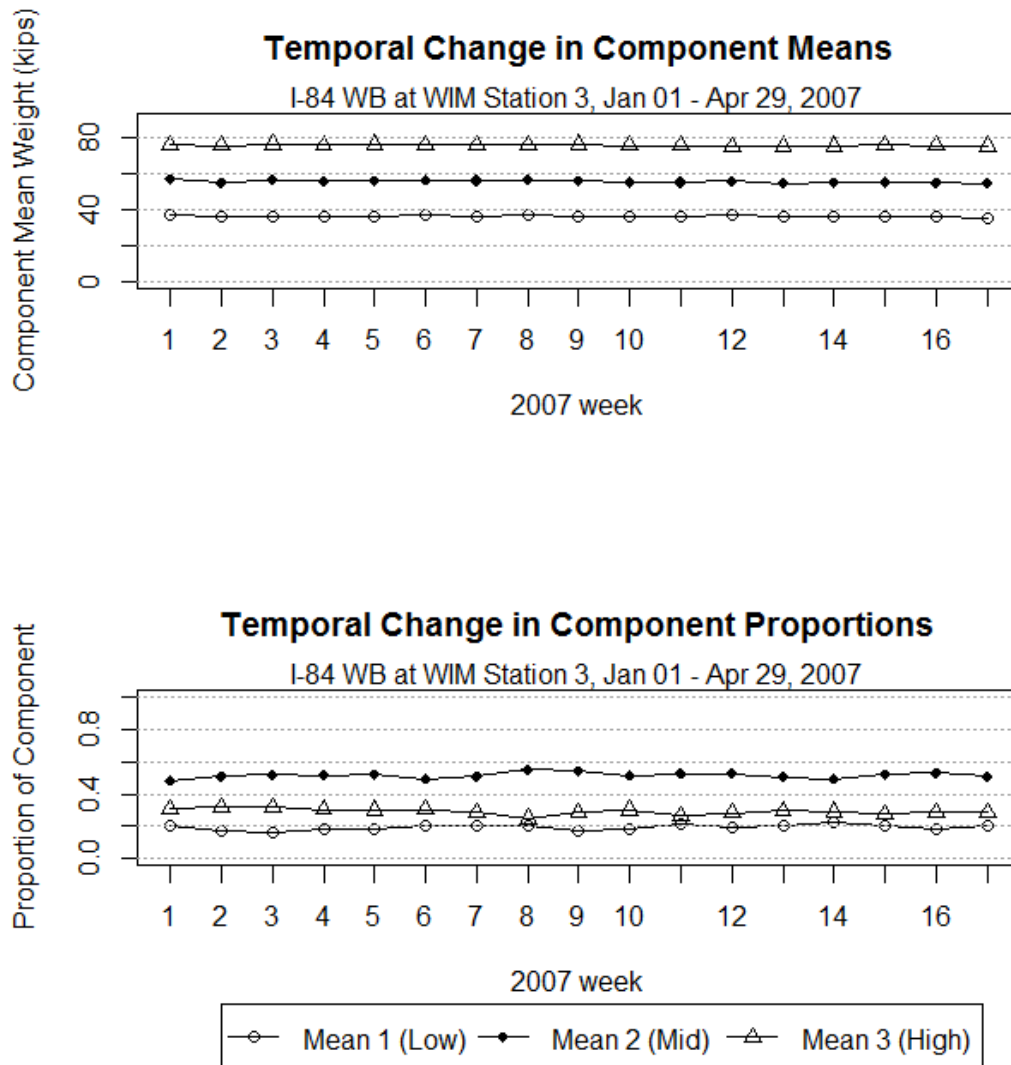


Figure 43: Component Mean and Proportions in WIM station 3

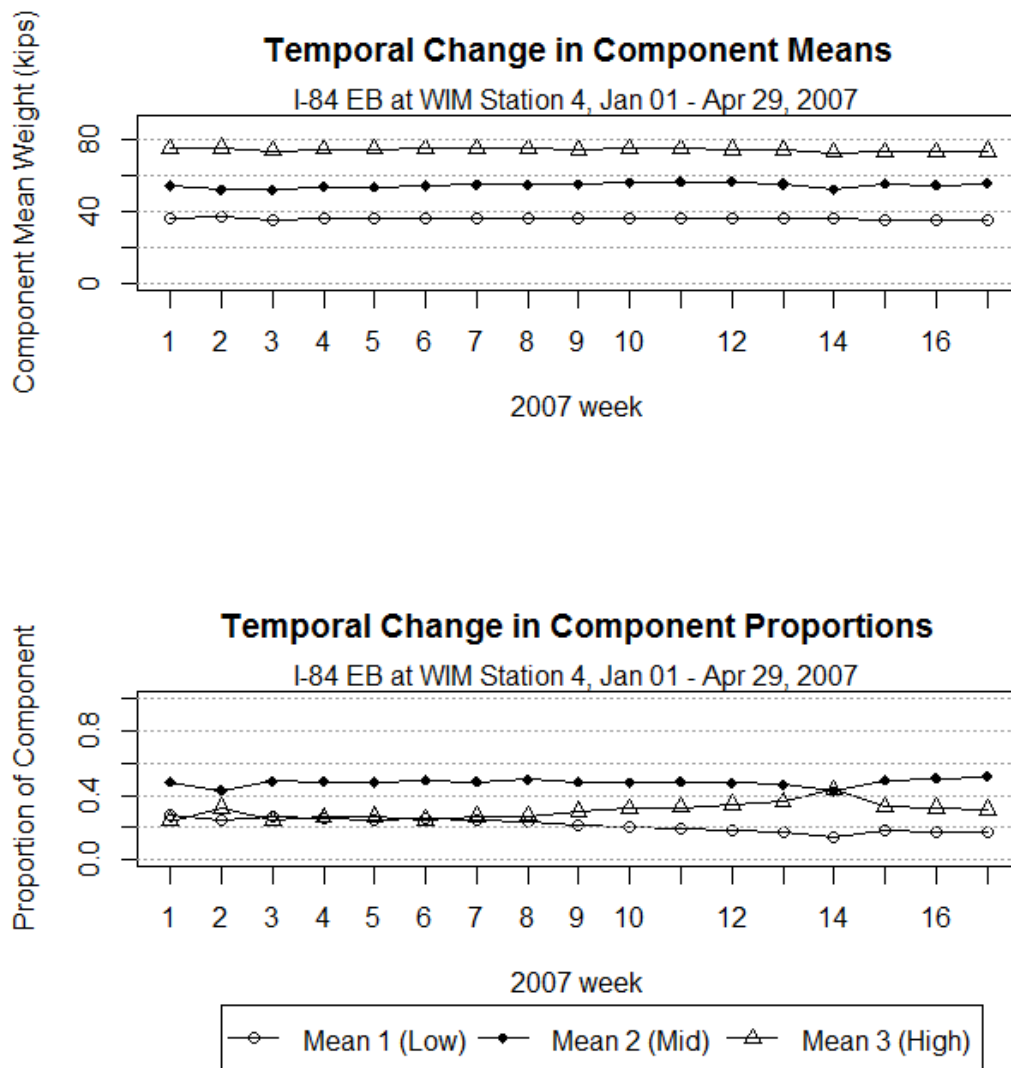


Figure 44: Component Mean and Proportions in WIM station 4

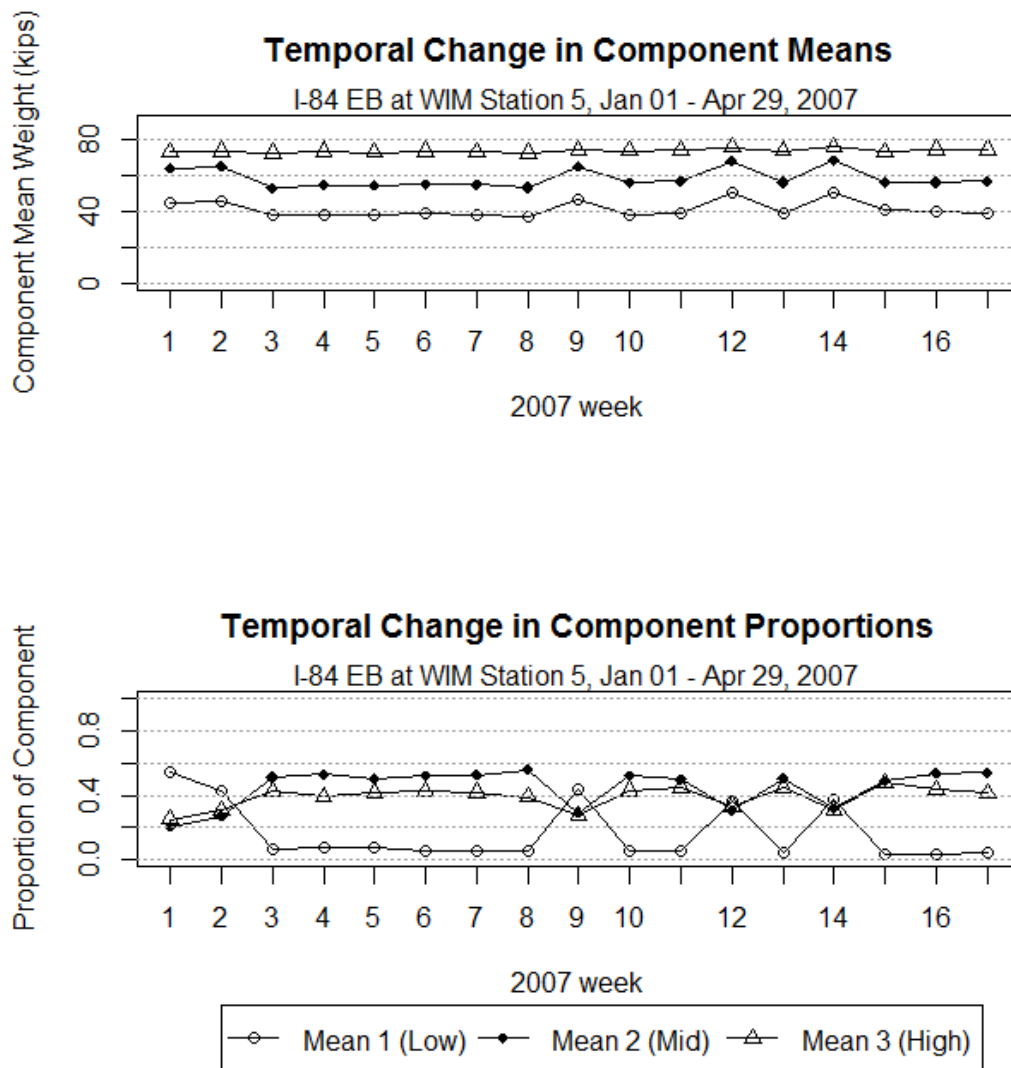


Figure 45: Component Mean and Proportions in WIM station 5

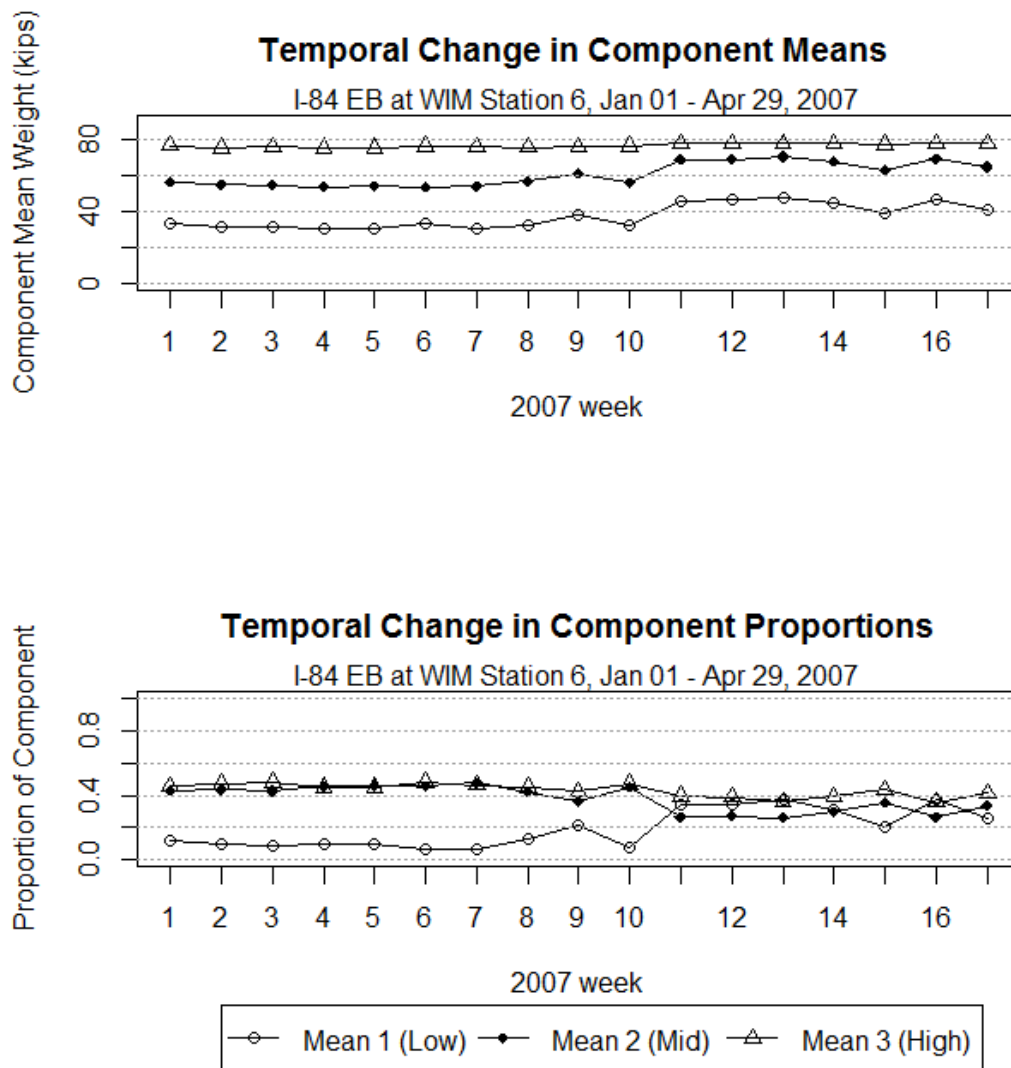


Figure 46: Component Mean and Proportions in WIM station 6

## APPENDIX D: Weekly Triangulation Plots

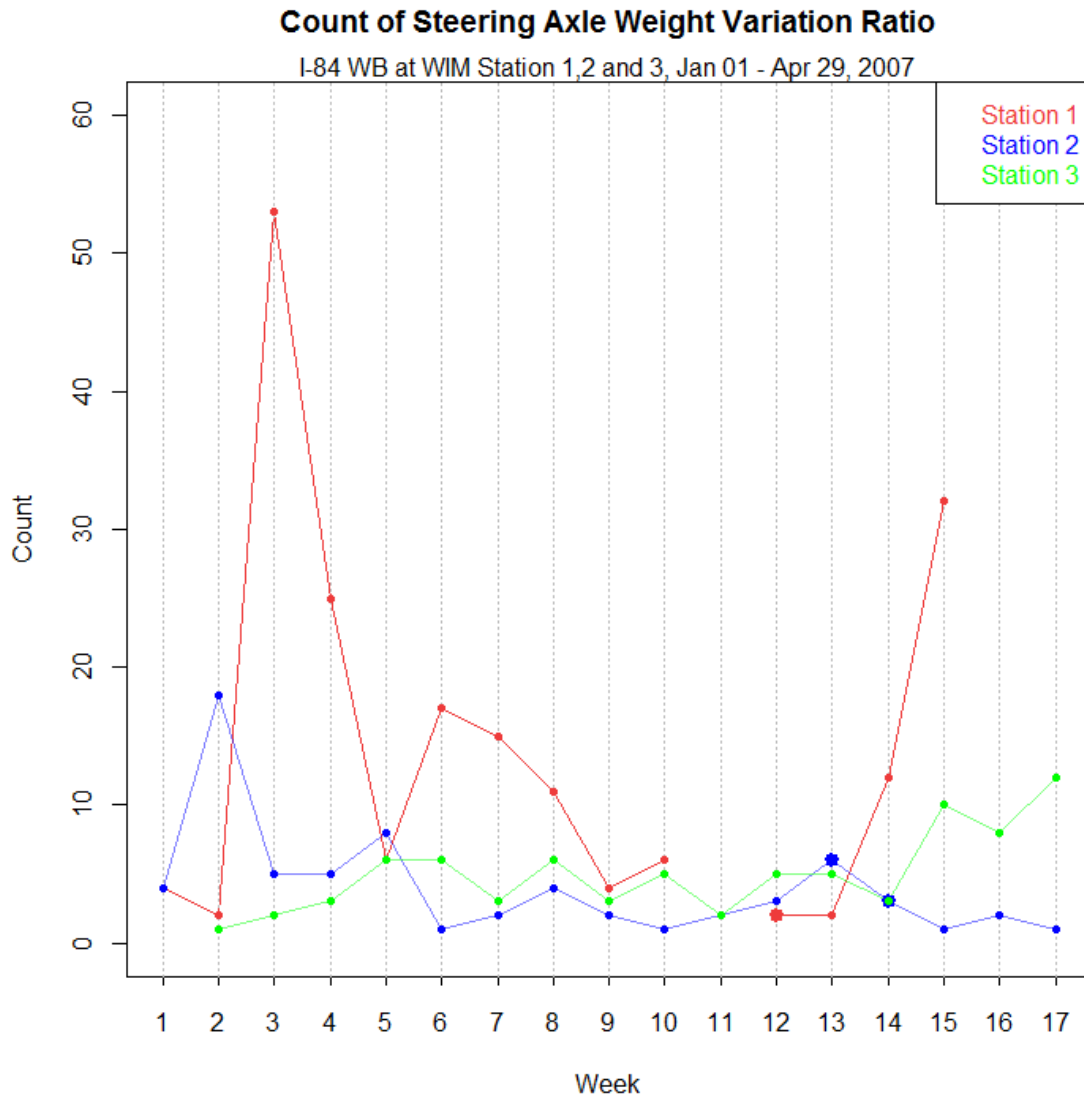
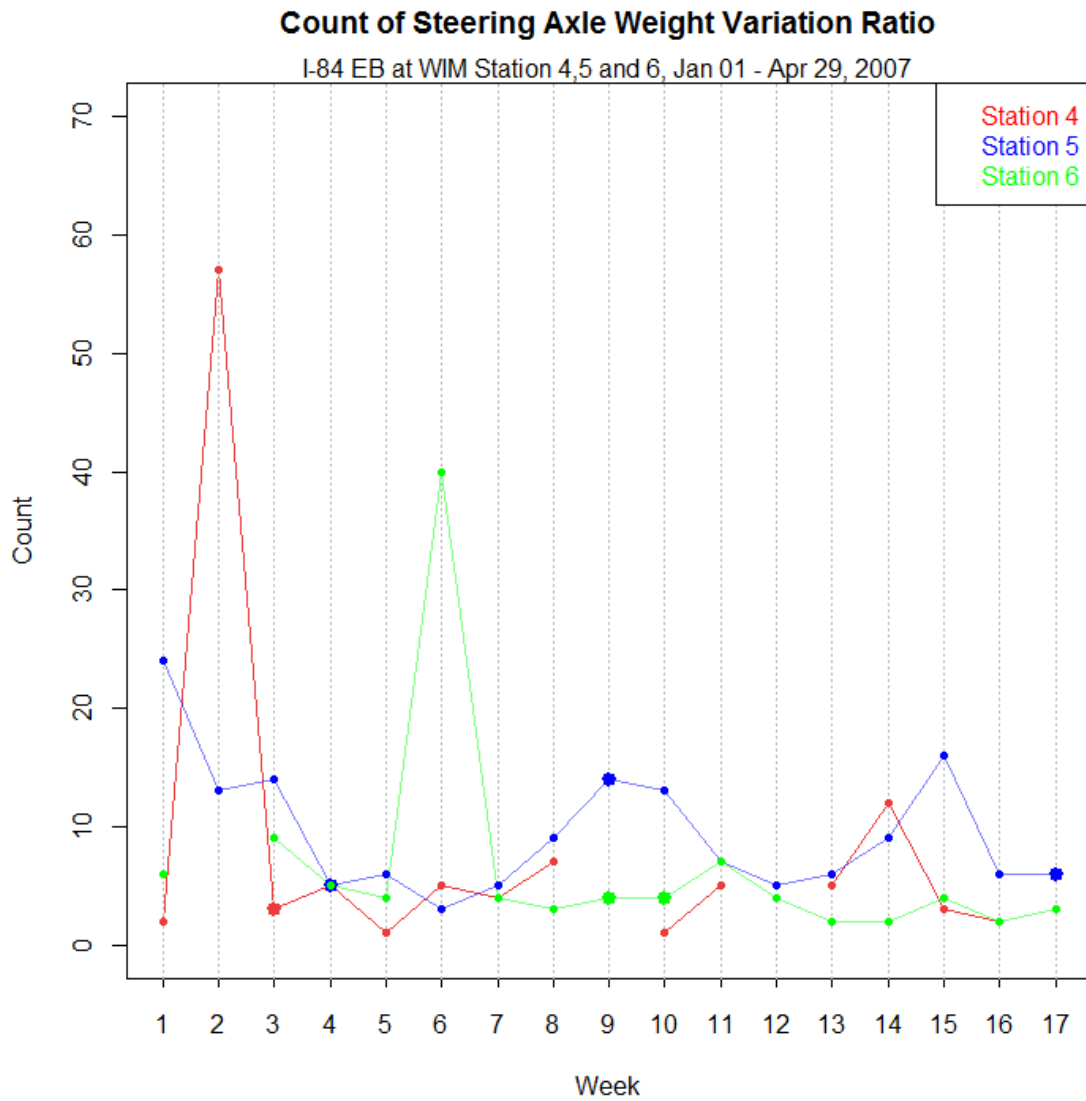
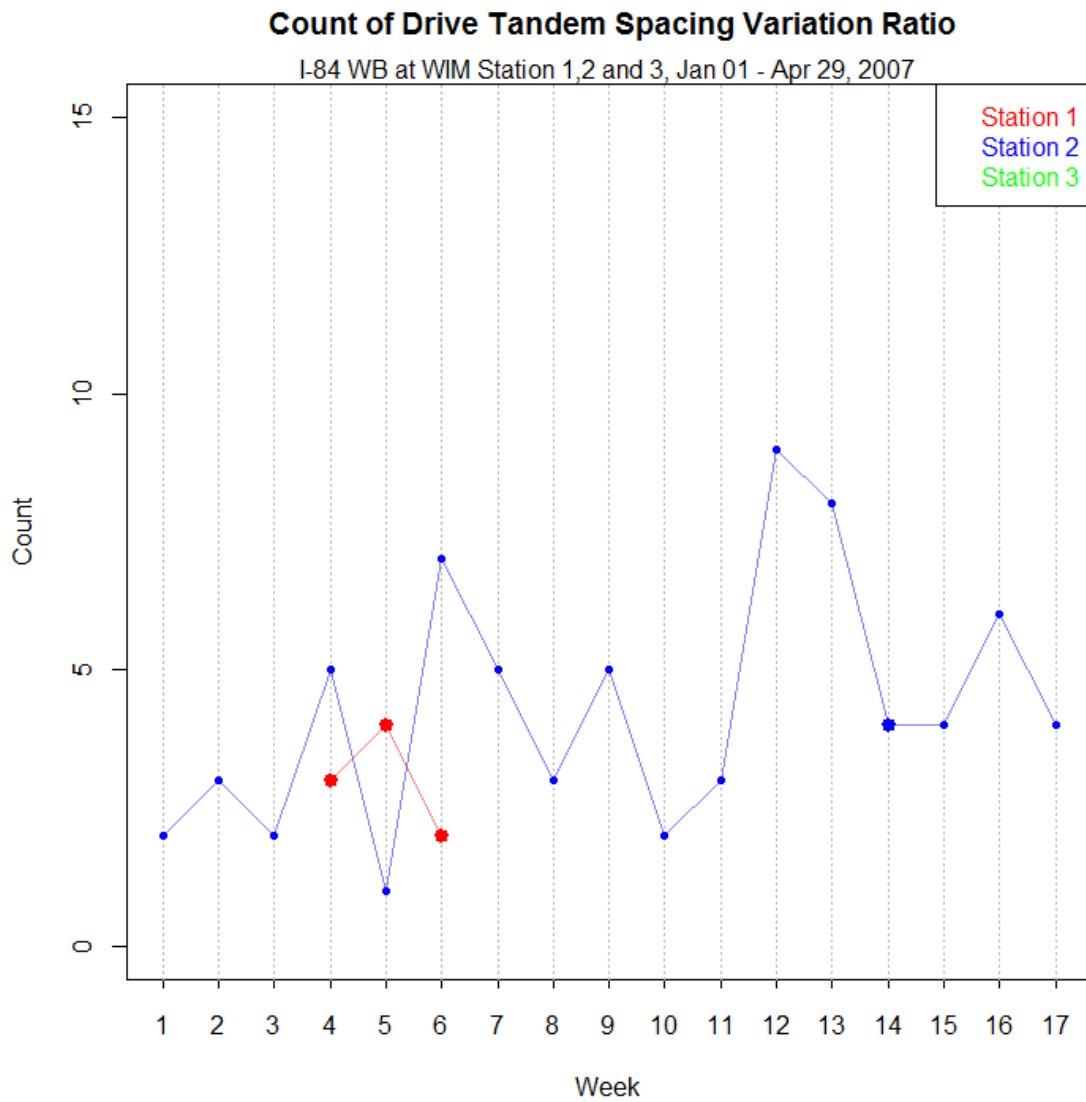


Figure 47: Weekly Triangulation Weight Plots at Station 1, 2 and 3

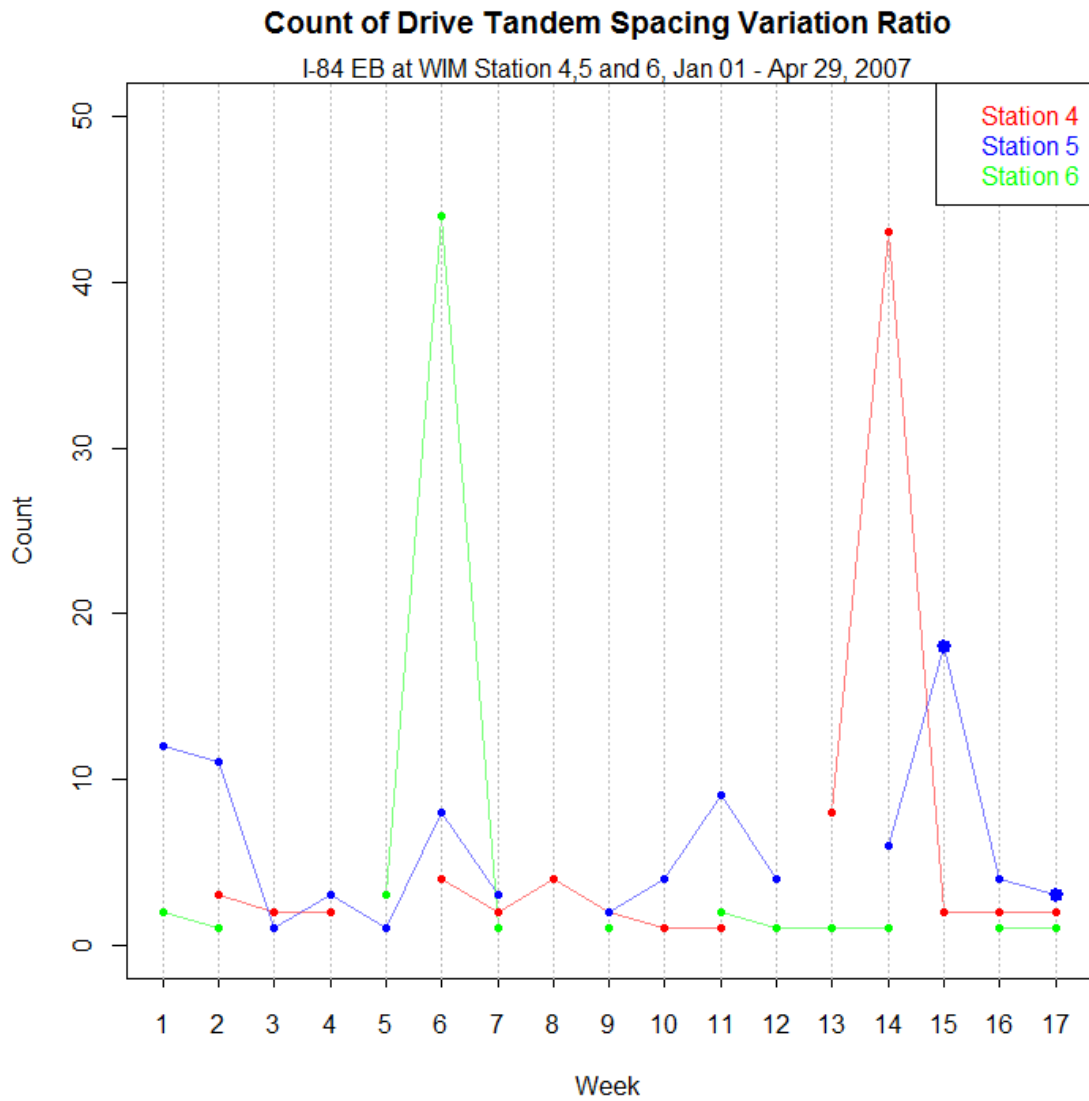


**Figure 48: Weekly Triangulation Weight Plots at Station 4, 5 and 6**





**Figure 49: Weekly Triangulation Axle Spacing Plots at Station 1, 2 and 3**



**Figure 50: Weekly Triangulation Axle Spacing Plots at Station 4, 5 and 6**